



HAL
open science

Génétique et Génomique de la capacité immunitaire chez le porc : approches eQTL et étude de biomarqueurs sanguins

Tatiana Maroilley

► To cite this version:

Tatiana Maroilley. Génétique et Génomique de la capacité immunitaire chez le porc : approches eQTL et étude de biomarqueurs sanguins. Biologie animale. Université Paris Saclay (COMUE), 2017. Français. NNT : 2017SACLV097 . tel-01734367

HAL Id: tel-01734367

<https://theses.hal.science/tel-01734367>

Submitted on 14 Mar 2018

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.

Génétique et génomique de la capacité immunitaire chez le porc : approche eQTL et étude de biomarqueurs sanguins

Thèse de doctorat de l'Université Paris-Saclay préparée à
l'Université de Versailles Saint-Quentin en Yvelines

ÉCOLE DOCTORALE N°577 STRUCTURE ET DYNAMIQUE DES SYSTEMES VIVANTS (SDSV)
SPECIALITE DE DOCTORAT: SCIENCES DE LA VIE ET DE LA SANTE

Thèse présentée et soutenue à Jouy-en-Josas, le 21 décembre 2017 par

Tatiana Maroilley

Sous la supervision de Claire Rogel-Gaillard et Jordi Estellé

Composition du Jury :

Henri-Jean Garchon PU-PH, Inserm	Président de jury
Sandrine Lagarrigue PhD, Agrocampus Ouest	Rapporteur
Gilles Fourcras PhD, INRA/ENVT	Rapporteur
Marie-José Mercat IE, IFIP/Bioporc	Examinateur
Juan J. Garrido PhD, Universidad de Cordoba	Examinateur
Claire Rogel-Gaillard PhD, INRA/AgroParisTech	Directrice de thèse

Table des matières

Table des figures.....	3
Table des tableaux.....	4
Remerciements.....	5
Liste des abréviations	6
Introduction.....	8
A. Le porc, animal de production : vers des systèmes durables qui incluent des caractères de santé.....	8
1. Quelques chiffres.....	8
2. Une filière sous contraintes.....	9
3. Organisation de la production française	10
4. La santé en élevage	15
B. Caractérisation de la compétence immunitaire	20
1. Contexte et définition.....	20
2. Architecture génétique des paramètres immunitaires	33
C. Cartographie du contrôle génétique de phénotypes	36
1 Le génome du porc et les puces de génotypage disponibles	37
2 Etude d'associations génome entier (GWAS).....	37
2. La génétique génomique ou cartographie du contrôle génétique du transcriptome	41
3. Lier « expression de gènes-phénotypes » ou identification de biomarqueurs	54
4. Génotype-transcriptome-phénotype	56
Les objectifs de la thèse	58
I. Etude 1 : Cartographie du contrôle génétique du transcriptome sanguin	60
II. Etude 2 : Etude des profils transcriptomiques de quatre tissus immunitaires	128
III. Etude 3 : Intégration de profils transcriptomiques sanguins et phénotypes immunitaires pour l'identification de biomarqueurs sanguins et de marqueurs génétiques pour des caractères immunitaires	
A. Introduction.....	174
B. Matériels et Méthodes	176
1. Animaux et échantillonnage.....	176
2. Extraction d'ARN et analyse du transcriptome	176
3. Phénotypage des traits immunitaires	177
4. Analyses génétiques et identification de biomarqueurs.....	186
C. Résultats	189
1. Variabilité des phénotypes immunitaires mesurés	189



2. Facteurs biologiques et d'environnement qui influent sur la variabilité des phénotypes immunitaires mesurés	191
3. Covariations des caractères immunitaires sanguins	193
4. Héritabilité des caractères immunitaires sanguins	199
5. Identification de biomarqueurs à partir du transcriptome sanguin.....	199
6. Identification de biomarqueurs candidats	203
7. Identification de marqueurs génétiques	206
D. Discussion	207
1. Une variabilité des paramètres immunitaires héréditaires, favorable pour leur prise en compte dans des programmes de sélection	207
2. L'immunophénotypage : une combinatoire complexe de mesures à simplifier ?	208
3. Vers l'identification de marqueurs génétiques et de biomarqueurs prédictifs des réponses immunes	209
Discussion générale	213
A. Les phénotypes immunitaires, le transcriptome du sang et son contrôle génétique comme ressources pour la recherche chez le porc	213
1. Etablissement de nouvelles ressources chez le porc	213
2. Intérêt pour les communautés scientifiques en production animale et dans le domaine biomédical 215	
3. Des ressources et des résultats à valider et à approfondir	216
B. Choix méthodologiques et limites des approches	218
1. Etude du transcriptome du sang	218
2. Identification de marqueurs génétiques de l'immunité en deux étapes	220
C. Perspectives en élevage	224
1. Confirmer l'utilisation de caractères immunitaires sanguins pour qualifier la compétence immunitaire individuelle	224
2. D'autres sources de biomarqueurs non invasives : le microbiote	226
3. Sélectionner les animaux sur leur compétence immunitaire.....	227
4. <i>Influence de l'environnement</i>	228
5. Vers une agriculture de précision qui intègre de multiples mesures.....	229
6. Edition des génomes ?.....	229
Bibliographie.....	231
Annexe 1 : Les projets ANR IMMOPIG et SUS_FLORA.....	244
Annexe 2 : Caractères immunitaires porcins mesurables dans le sang	247
Publications et Communications	251



Table des figures

Figure I-1 : Evolution de la production et de la consommation en France	9
Figure I-2 : Structure pyramidale à trois niveaux de la filière porcine	11
Figure I-3 : Répartition du poids de chaque critère de sélection dans les lignées Large White françaises mâle et femelle.....	12
Figure I-4 : Principe de la sélection génomique	15
Figure I-5 : Organisation temporelle des réponses immunitaires innées et adaptatives	21
Figure I-6 : Organes lymphoïdes du porc	22
Figure I-7 : Origine et différenciation des cellules immunitaires	24
Figure I-8 : Schéma simplifié représentant la structure tissulaire d'une plaque de Peyer	25
Figure I-9 : Présentation des peptides par les molécules du complexe majeur d'histocompatibilité (CMH) qui assurent la reconnaissance du soi et du non soi	26
Figure I-10 : Mise en place de la réaction immunitaire	28
Figure I-11 : Représentation de l'implication du sang dans le système immunitaire	29
Figure I-12 : Ensemble des techniques applicables sur le sang pour phénotyper l'immunité.....	30
Figure I-13 : Caractères immunitaires sanguins mesurés par Flori <i>et al.</i> (2011) et héritabilités estimées.....	35
Figure I-14 : Analyse d'association génome entier	38
Figure I-15 : Nombre de QTL associés à des traits phénotypiques	40
Figure I-16 : Représentation canonique du processus ADN-ARN-protéine	43
Figure I-17 : Organisation génomique des transcrits codants et non codants	44
Figure I-18 : Protocole d'utilisation des puces à expression	45
Figure I-19 : Protocole de séquençage d'ARN	46
Figure I-20 : Protocole d'analyse standard de données RNA-Seq.....	48
Figure I-21 : Mécanismes moléculaire des variants régulateurs <i>cis</i> /local et <i>trans</i> /distant	50
Figure I-22 : Représentation d'effet ASE détecté par séquençage	53
Figure O-23 : Représentation des étapes d'intégrations des trois niveaux biologiques abordés.....	58
Figure O-24 : Représentation des objectifs spécifiques du travail de thèse dans les cadres à fond violet. ...	59
Figure E1-25 : Workflow d'analyse suivi pour l'eGWAS	60
Figure E1-26 : Workflow d'analyse suivi pour l'analyse ASE	61
Figure E2-27 : Workflow d'analyses suivi pour l'étude 2	128
Figure E3-28 : Stratégie d'analyse des données de cytométrie en flux pour analyser la capacité de phagocytose.	179
Figure E3-29 : Stratégie d'analyse des données de cytométrie en flux pour identifier les cellules NK et NKT dans les PBMC du sang de porc.	181
Figure E3-30 : Stratégie d'analyse des données de cytométrie en flux pour identifier les lymphocytes T $\gamma\delta$, les lymphocytes T cytotoxiques (T CD4-CD8+), les lymphocytes T helper CD4 (CD4+ CD8-) et les lymphocytes T mémoire (T CD4+CD8+) dans les PBMC du sang de porc.	182
Figure E3-31 : Stratégie d'analyse des données de cytométrie en flux pour identifier les cellules dendritiques (cDC1 et cDC2), les 4 types de monocytes (I à IV désignant des stades de maturation croissants) dans les PBMC du sang de porc..	183
Figure E3-32 : Schéma descriptif des analyses statistiques effectuées	186
Figure E3-33 : Matrice des corrélations (Spearman) entre les 55 caractères immunitaires (Corrplot).....	194



Figure E3-34 : Classification hiérarchique non supervisée répartissant les caractères immunitaires en 8 groupes.....	196
Figure E3-35 : Distribution des coefficient de corrélations.....	197
Figure E3-36 : Représentation du réseau des corrélations directes significatives entre les caractères immunitaires, déterminées par PCIT.....	198
Figure E3-37 : Estimation de l'héritabilité et de l'écart-type correspondant pour 18 caractères immunitaires	199
Figure E3-38 : sPLS pour les caractères de phagocytose	204
Figure E3-39 : sPLSA pour une partie des caractères du cluster 1	205
Figure A-40 : A) Protocole de production des animaux du projet global ; B) Etapes de production des animaux élevés au Rheu et utilisés dans cette thèse.....	244
Figure A-41 : A) Animaux produits dans le projet SUS_FLORA; B) Protocole de production des animaux du projet SUS_FLORA	246

Table des tableaux

Tableau I-1 : Evolution de la production et de la consommation en France	8
Tableau I-2 : Principales caractéristiques des lignées sélectionnées par race.....	13
Tableau I-3 : Classement et caractéristiques des principales maladies porcines	18
Tableau I-4 : Organisation et fonctionnalités des organes lymphoïdes primaires et secondaires	23
Tableau I-5 : Principales cellules de l'immunité innée	25
Tableau I-6 : Sélection d'études eGWAS publiées chez le porc	49
Tableau E3-7 : Liste des anticorps et cytokines recombinantes utilisés dans le test de dosage des cytokines porcines en multiplex. Les limites de quantification constatées avec ce test sont indiquées, sans tenir compte du facteur de dilution des échantillons.....	184
Tableau E3-8 : Résumé des facteurs pris en compte dans le modèle pour estimer l'héritabilité par caractère immunitaire	187
Tableau E3-9 : Statistiques générales des caractères immunitaires sanguins étudiés	189
Tableau E3-10 : Effets qui influent de manière significative sur les caractères immunitaires (code : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1)	191
Tableau E3-11 : Nombre de sondes et gènes dont le niveau d'expression a été trouvé associé à la variation des phénotypes	200
Tableau E3-12 : Top 20 des fonctions biologiques trouvées enrichies sur la base les gènes dont la variation d'expression dans les sang a été trouvée associée à la variation d'au moins un paramètre immunitaire descriptif ou de réponse.....	201



Remerciements

Cette thèse a été réalisée au sein de l'équipe Génétique Immunité Santé (GIS) de l'unité Génétique Animale et Biologie Intégrative (GABI) de l'INRA de Jouy-en-Josas. Le contrat de thèse a lui été co-financé par l'association BIOPORC, le regroupement des entreprises françaises de sélections porcines et le département de Génétique Animale de l'INRA.

Mes premiers remerciements vont au Dr Claire Rogel-Gaillard, Directrice de l'Unité GABI et directrice de cette thèse pour son accueil au sein de son laboratoire, pour son accompagnement tout au long de mon projet, pour m'avoir appris pendant trois ans à voir le travail de recherche dans sa globalité, à se projeter toujours plus loin et à toujours chercher la précision. Je remercie le Dr Jordi Estellé, co-encadrant de cette thèse pour son aide et son soutien.

Je tiens également à remercier mes deux encadrants pour m'avoir permis de m'investir autant dans le monde de la recherche, tant par la participation à des congrès, que par mon implication au sein de l'association Doc'J des doctorants du centre INRA de Jouy-en-Josas et au sein de l'Ecole Doctorale SDSV en tant que représentante des doctorants et enfin par la mission doctorale réalisée au cours de ma deuxième année de thèse au sein de la Cellule Europe de l'Université Paris-Saclay.

Je remercie très chaleureusement Bioporc pour le co-financement de la thèse et tous ces membres d'avoir pris le temps de me rencontrer à plusieurs reprises afin de suivre mon travail de thèse. Je tiens à remercier plus particulièrement Marie-José Mercat, pour sa disponibilité et Loïc Flatres-Grall pour le temps qu'il a consacré à me faire découvrir la sélection porcine sur le terrain.

Je remercie par ailleurs les différents membres du jury et notamment les Dr Sandrine Laguarrigue et Gilles Foucras d'avoir accepté d'être rapporteurs de ce travail de thèse. Je remercie également le Dr Juan J. Guarrido, Marie-José Mercat et le Dr Henri-Jean Garchon.

Je remercie également les membres de mon comité de thèse, les Dr Catherine Larzul, Yulixis Ramayo, Henri-Jean Garchon et Guillaume Laval ainsi que Marie-José Mercat et Loïc Flatres-Grall pour le temps qu'ils m'ont accordé afin de me conseiller durant ce travail de thèse.

Je tiens à remercier tous mes formidables collègues du bâtiment 320, tous les autres membres de l'unité GABI et merci à tous ces jeunes chercheurs que j'ai pu rencontré grâce à l'association Doc'J.

Enfin, je remercie ma famille et tout particulièrement mon mari pour son soutien indéfectible.



Liste des abréviations

ADN : Acide Désoxyribo Nucléique

ARN : Acide Ribo-Nucléique

ARNInc : ARN long non codant

ARNm : ARN messenger

ASE : Expression allèle spécifique (*Allele Specific Expression*)

CCMH : Concentration Corpusculaire Moyenne en Hémoglobine

cDC1/2 : Cellules Dendritiques de type 1

CEA : Centre d'Energie Atomique

CLR : *C-type lectin like Receptors*

CMH : Complexe Majeur d'Histocompatibilité

DE : Différentiellement Exprimé

DL : Déséquilibre de Liaison

eGWAS : Etude d'association d'expression génome entier (*expression Genome-Wide Association Study*)

eQTL : *expression Quantitative Trait Loci*

FDR : Taux de faux positifs (*False Discovery Rate*)

$\gamma\delta$ T : Lymphocytes T $\gamma\delta$

GWAS : Etude d'association génome entier (*Genome-Wide Association Study*)

IFIP : Institut du Porc

IL2/8/10/12/17/1 β : Interleukine 2/8/10/12/17/1 β

INF γ : Interféron γ

INRA : Institut National de la Recherche Agronomique

IPA : Ingenuity Pathway Analysis®

IPP : Plaque de Peyer iléale (*Ileal Peyer Patches*)

JPP : Plaque de Peyer jéjunale (*Jejunal Peyer Patches*)

LPS : Lipopolysaccharide

LW : Large White

MLN : Ganglion mésentérique (*Mesenteric Lymph Node*)

NFS : Numération de Formule Sanguine

OP : Organisation de Producteur

OSP : Organisme de Sélection Procine

PB : Sang périphérique (*Peripheral Blood*)



PBMC : Cellules mononucléées périphériques sanguines (*Peripheral Blood Mononucleic Cells*)

PCIT : Partial Coefficient with Information Theory

PEA : Entérotype porcin A (*Porcine Enterotype A*)

PEB : Entérotype porcin B (*Porcine Enterotype B*)

PI : PMA-ionomycine

PIL : PMA-ionomycine-Lipopolysaccharide

QTL : Locus associé à un trait quantitative (*Quantitative Trait Loci*)

RNA-Seq : Séquençage d'Acide Ribo-Nucléique

RIN : *RNA Integrity Number*

RT-PCR : Réaction de Polymérisation en Chaîne après Transcription Inverse (*Reverse Transcription and Polymerase Chain Reaction*)

SLA : Complexe Majeur d'Histocompatibilité du porc (*Swine Leucocyte Antigen*)

SNP : Polymorphisme Nucléotidique (*Single Nucleotide Polymorphism*)

TCMH : Teneur Corpusculaire Moyenne en Hémoglobine

TNF α : Facteur de Nécrose Tumorale (*Tumor Necrosis Factor α*)

VG : Valeur Génétique

VGM : Volume Globulaire Moyen



Introduction

A. Le porc, animal de production : vers des systèmes durables qui incluent des caractères de santé

En France, l'institut du porc ou IFIP est une structure de recherche-développement-innovation au service de tous les métiers de la filière porcine. Son expertise facilite l'optimisation des réponses face aux défis rencontrés par la filière. L'IFIP est également un observateur des résultats de la filière porcine et les chiffres décrits dans cette partie sont pour la plupart issus de sa publication annuelle « Le porc par les chiffres 2016-2017 » ainsi que de son bilan d'activité 2016.

1. Quelques chiffres

Le porc est un animal de consommation issu d'une importante filière d'élevage. La production porcine mondiale est égale à 114 millions de tonnes dont la moitié réalisée en Chine (IFIP, Institut du Porc, 2016a). En 2015, la France a produit 2,23 millions de tonnes (Tableau I-1). Cette production est destinée à la fois à la consommation nationale et aux exportations qui représentent 26% de la production (580 000 tonnes). La France importe à hauteur de 25% de sa consommation (613 000 tonnes), pour satisfaire des besoins mal couverts par l'offre nationale, en termes d'élaboration, de qualité ou de spécifications des produits (jambon Ibérique ou Serrano). Les produits d'importation sont parfois plus compétitifs que les produits nationaux.

Tableau I-1 : Evolution de la production et de la consommation en France (source : Le porc en chiffres 2015-2016, IFIP Institut du porc)

en milliers de tonnes équivalent carcasses	2013	2014	2015*	2015/14 (%)
Production (PIB) ⁽¹⁾	2 206	2 209	2 228	+0,9%
Importations	600	595	613	+3,0%
Exportations	668	610	578	-5,3%
Consommation (CIB) ⁽¹⁾	2 138	2 194	2 264	+3,2%
Solde extérieur	68	15	-35	-
Auto-provisionnement	1 03,2	100,7	98,4	-2,2%
Consommation par kg/habitant	32,5	33,2	34,0	+2,7%

*Données 2015 : estimations Ifip ; (1) PIB : Production Indigène Brute. CIB : Consommation Indigène Brute. Y compris autoconsommation. Carcasses avec tête, sans panne, rognons, ni diaphragme. Ensemble Métropole + DOM.

Source : IFIP d'après SSP-AGRESTE, EUROSTAT, DOUANES.



Historiquement, la croissance de la production porcine française a été très dynamique dans les années 1980-1990. Mais, à la fin des années 1990, les mesures de protection de l'environnement ont ralenti cette croissance. La production porcine s'est diversifiée : produits de charcuterie, produits cuits, produits de salaisons sèches et saumurées, produits de traiteurs. Depuis quelques années se développe également la filière biologique, ralentie cependant par le manque de matières premières riches en protéines (Alibert *et al.*, 2016). Toutefois, les productions de porcs sous Label Rouge et issues de l'Agriculture Biologique totalisent près de 4 % du total français. La production porcine en France est aujourd'hui organisée autour de 37 groupements de producteurs ou organisations de producteurs (OP). Ces groupements commercialisent plus de 90 % des porcs charcutiers abattus en France. Ce sont, pour la plupart, des coopératives ou unions de coopératives.

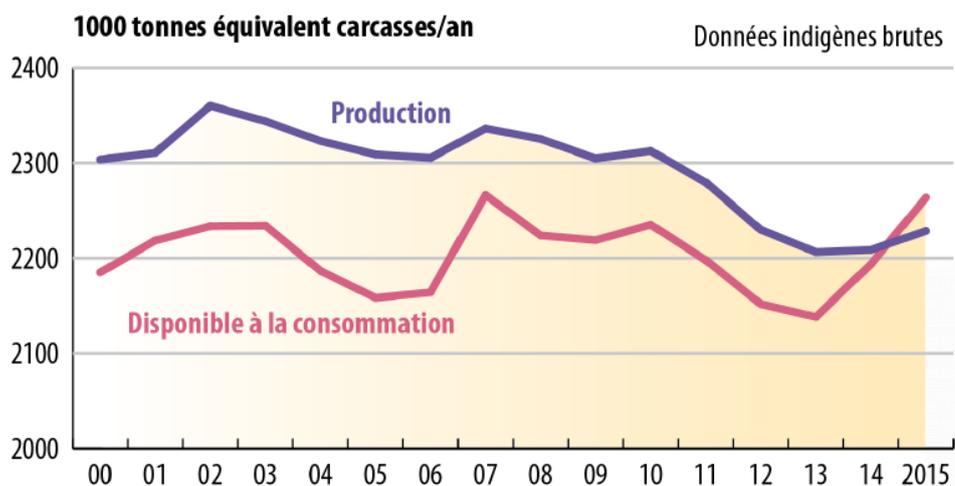


Figure I-1 : Evolution de la production et de la consommation en France (source : Le porc en chiffres 2015-2016, IFIP Institut du porc)

2. Une filière sous contraintes

La production porcine en France est soumise à de nombreuses contraintes et pressions et il est essentiel d'y faire face afin de maintenir une productivité suffisante permettant la rentabilité de la filière pour tous ses acteurs, des sélectionneurs aux producteurs. De plus, la filière se confronte actuellement à une baisse de la consommation de viande. En 2015, la consommation de porc a atteint 34 kg équivalent carcasses/habitant. En 2016, elle a reculé de 3% en volume pour le porc frais et de 0,5% pour la charcuterie (IFIP, Institut du Porc, 2016b). La filière porcine doit donc évoluer pour conjuguer rentabilité et durabilité, face à une baisse de consommation liée pour partie à des changements sociétaux.



Il est donc important de financer des projets permettant aux professionnels de la filière porcine de comprendre les attentes des consommateurs et d'adapter leurs choix de production. Un sondage a ainsi été réalisé en juin 2016 (projet ACCEPT ; Delanoue *et al.*, 2017), auprès de 2000 personnes, représentatives de la population française (panel IFOP) pour quantifier leur connaissance, leur intérêt et vision de l'élevage français, leur opinion sur diverses pratiques contraignantes pour les animaux, les objectifs qu'ils assignent à l'élevage et les services qu'ils en attendent, ainsi que leurs habitudes de consommation et les informations souhaitées sur les produits animaux. Les résultats de cette enquête montrent que 1,4% des personnes se déclarent végétariennes, 0,15% se qualifient végans, 14% pensent cesser leur consommation de viande dans les 12 prochains mois et 18% la réduire. Les personnes souhaitant cesser leur consommation sont plutôt des jeunes et des femmes. Deux tiers des personnes interrogées (dont 55% entre 18 et 24 ans) n'envisagent donc ni de cesser, ni de diminuer leur consommation de viande (Delanoue *et al.*, 2017). Ainsi, il est clair que si les citoyens français consomment moins de viande, ils recherchent en contrepartie plus de « qualité » selon différents critères : nutritionnels, sanitaires, gustatifs, éthiques, sociétaux, d'usage. Lors du sondage du projet ACCEPT, 59% des personnes enquêtées ne sont pas satisfaites des conditions de vie des animaux et de leur bien-être (dont 38% pas du tout satisfaites), 52% de l'échantillon ne sont pas satisfaits du respect de l'environnement par les éleveurs, 40% sont insatisfaits de la sécurité sanitaire des produits, et 33% de l'entretien des paysages par les pratiques d'élevage. L'accès des animaux au plein air et l'information du consommateur sur les modes d'élevage constituent les attentes les plus prégnantes.

L'évolution des systèmes de production et les recherches associées ont ainsi pour but d'appuyer la filière porcine face aux exigences de la multi-performance en élevage qui vise à la fois la productivité, la qualité des produits, le bien-être animal, ainsi que le respect de l'environnement et de l'attente des citoyens.

3. Organisation de la production française

a) Une structuration pyramidale

En France, l'Agence de la Sélection Porcine (ASP) est l'organe de représentation des professionnels de la génétique porcine pour la sélection. Elle traite des dossiers techniques à la demande de ses adhérents ou du Ministère chargé de l'Agriculture. La sélection porcine collective est organisée autour d'Organismes de Sélection Porcine (OSP), dont la plupart se sont regroupés en une association : Bioporc. Les OSP coordonnent et orientent l'amélioration génétique.



La production porcine en France est organisée selon un système pyramidal à 3 étages : sélection, multiplication et production (Figure I-2). A son sommet, les OSP ont pour but d'améliorer le niveau de performances moyennes des populations porcines sur des caractères d'intérêt économique, en exploitant des schémas de croisement, de façon à combiner les aptitudes particulières de chaque type génétique (Tableau I-2). Ce travail d'amélioration génétique consiste à repérer les meilleurs individus d'une génération, sur la base de l'estimation de leur valeur génétique (VG) pour les objectifs de sélection définis. Ces individus seront ensuite gardés comme reproducteurs et leur semence diffusée par des centres d'insémination artificielle (CIA). Les élevages de multiplication sont ensuite chargés de produire les truies parentales et les verrats terminaux. La semence des verrats et les truies sont par la suite vendues aux élevages de production pour l'élevage des porcs charcutiers commercialisés.

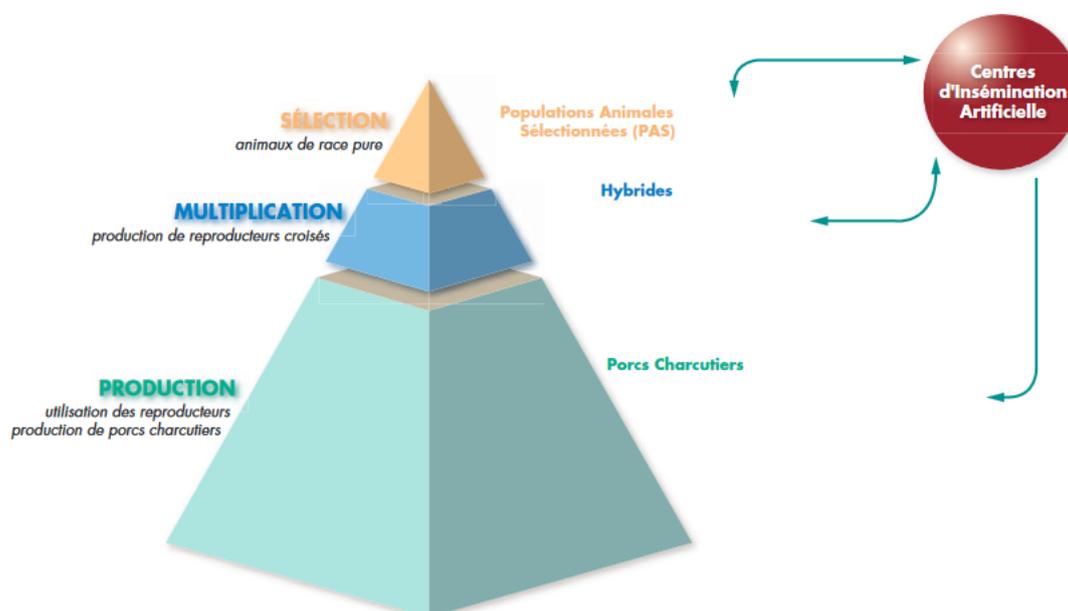


Figure I-2 : Structure pyramidale à trois niveaux de la filière porcine (source : La Sélection porcine collective en France 2003, ITP)

Des modèles statistiques permettent de prédire la VG des candidats à la sélection à partir de leurs performances propres et de celles de leurs apparentés et contemporains. Ainsi, en France, chaque semaine, cinq populations porcines sont évaluées sur des critères phénotypiques et les VG sont transmises aux sélectionneurs, OSP, groupements d'éleveurs et CIA. Parmi elles, quatre populations sont collectives aux OSP adhérentes de Bioporc (Large White lignée Femelle, Landrace, Piétrain et Large White lignée Mâle) et la dernière est une population sélectionnée autonome (Duroc Axiom). Les performances des animaux enregistrées en élevage de sélection, de multiplication ou en station de contrôle sont centralisées dans la base nationale de données génétiques porcines « Banapog ». Seules les données validées sont conservées



dans la base de données pour l'indexation et la recherche (BDir), qui regroupe fin 2016 plus de 11 millions d'individus avec des données valides (IFIP, Institut du Porc, 2016b).

b) Des objectifs de sélection qui évoluent

Pendant longtemps, la sélection s'est faite sur l'observation de caractères d'importance économique facilement mesurables comme le poids, la croissance, la conformation de carcasse, le gras intramusculaire, l'épaisseur de lard dorsal. Les porcs d'aujourd'hui sont ainsi le fruit d'une sélection de plus de 30 ans pour le gain de poids et une viande maigre. La sélection s'est progressivement élargie à d'autres caractères, comme la prolificité des truies, l'efficacité alimentaire (Sanchez *et al.*, 2014; Tribout *et al.*, 2013).

Le Tableau I-2 résume les caractères sur lesquels chaque race/lignée est sélectionnée. Les lignées femelles et les lignées mâles sont sélectionnées sur des critères différents. Les principaux axes de sélection des lignées femelles portent sur la productivité numérique des truies, leurs qualités maternelles et les caractères de production. Les principaux axes de sélection des lignées mâles portent sur la croissance des animaux en engraissement, l'efficacité alimentaire, la composition des carcasses et la qualité de la viande. La Figure I-3 représente le poids de chaque caractère dans la sélection collective actuelle des OSP pour les lignées mâles et femelles Large White françaises.

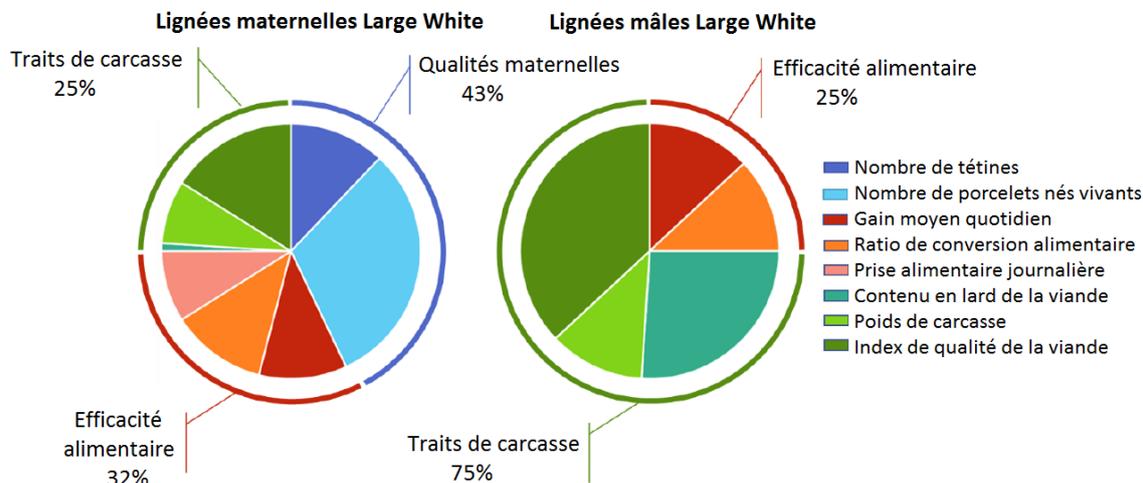


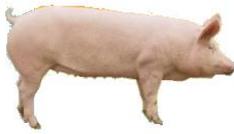
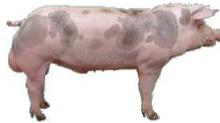
Figure I-3 : Répartition du poids de chaque critère de sélection dans les lignées Large White françaises mâle et femelle.

Les objectifs de sélection s'étendent vers des caractères liés au bien-être. Ainsi, des recherches sont conduites sur l'autonomie et l'aptitude maternelle des truies pour améliorer le taux de survie des porcelets à la naissance. Des travaux sont également menés sur l'identification d'alternatives à la castration des mâles. En complément, dans un contexte de changement global, réduire l'empreinte environnementale des



élevages et préserver les ressources devient une nécessité, d’où la mise en place de nombreux projets relatifs à la sobriété : produire mieux et différemment. Parmi les caractères étudiés , la résistance aux pathogènes occupe une place importante, notamment pour limiter l’usage des anti-infectieux (antibiotiques notamment). Une analyse de la part génétique de ces phénotypes liés à la santé devrait permettre d’introduire ces nouveaux objectifs de sélection dans les schémas d’amélioration génétique, renforçant ainsi le contrôle sanitaire et le bien-être en élevage.

Tableau I-2 : Principales caractéristiques des lignées sélectionnées par race en France

Lignées femelles	<i>Large White femelle</i>		Grande résistance et facilité d'adaptation Excellente prolificité Bonne vitesse de croissance Bon indice de consommation Bonne qualité de viande
	<i>Landrace Français</i>		Bonne prolificité Bonnes qualités maternelles Bonnes performances d'engraissement Bonne qualité de viande
	<i>Sino-européennes</i>		Développées à partir des années 1980 Allient la prolificité et les qualités maternelles des races chinoises et les performances de croissance et de composition corporelle des races européennes
	<i>Duroc femelle</i>		Rusticité et robustesse Excellente qualité de viande Bonnes performances de croissance Souvent utilisées en croisement 3 voies
Lignées mâles	<i>Piértrain</i>		Excellent indice de consommation Taux de muscle exceptionnel Prise en compte du génotpe Halothane Bonne conformation
	<i>Large White mâle</i>		Croissance exceptionnelle Bon taux de muscle Bonne qualité de viande
	<i>Duroc mâle</i>		Rusticité et robustesse Bonnes performances de croissance Bon indice de consommation Réputé pour la qualité de la viande et le gras intramusculaire
	<i>Synthétiques</i>		Crées à partir des principales races, elles allient : Excellentes performances de croissance Bon indice de consommation Taux de muscle élevé Bonne qualité de viande

c) Potentiel de la sélection génomique

Depuis peu, la sélection génomique permet la prédiction de la valeur génétique d'un individu à partir des marqueurs couvrant le génome, avant la mesure des performances, lorsqu'une population de référence, à la fois génotypée et phénotypée peut être constituée (Figure I-4). Cette approche a révolutionné la génétique bovine (Boichard, 2015; Boichard *et al.*, 2016) car elle permet la sélection des individus sur la base de leur génotype, dès leur plus jeune âge.

La sélection génomique se met progressivement en place chez le porc en France, avec des étapes pour estimer les gains par rapport à l'existant (Tribout *et al.*, 2012, 2013). En effet, les bovins sont sélectionnés et produits en races pures alors que les porcs sont sélectionnés en races pures mais produits en croisement. Selon les résultats du programme Utopige (projet ANR, coordination Pascale Leroy, INRA) pour une Utilisation Optimale de l'Information Génomique dans les schémas Pyramidaux, la sélection génomique permet une amélioration de 30% du progrès génétique (notamment pour le critère du nombre de porcelets nés vivants) et elle est rentable en seulement quelques années.

La sélection génomique permet l'évaluation en une seule étape des phénotypes, génotypes et du pedigree des animaux. Elle permet d'évaluer des critères faiblement héritables. Elle est efficace sur les races pures comme les croisés. En 2016, la sélection génomique a été déployée dans la population Landrace français avec une puce de marqueurs génétiques à moyenne densité (60K). L'IFIP développe actuellement une chaîne d'évaluation génomique dans la population Large White lignée femelle et la diffusion des premiers index génomiques a eu lieu au cours du 1er semestre 2017.

Le déploiement de la sélection génomique en porc devrait être un atout pour inclure des caractères complexe, à faible déterminisme génétique mais forte plus-value potentielle en élevage comme probablement les caractères de santé.



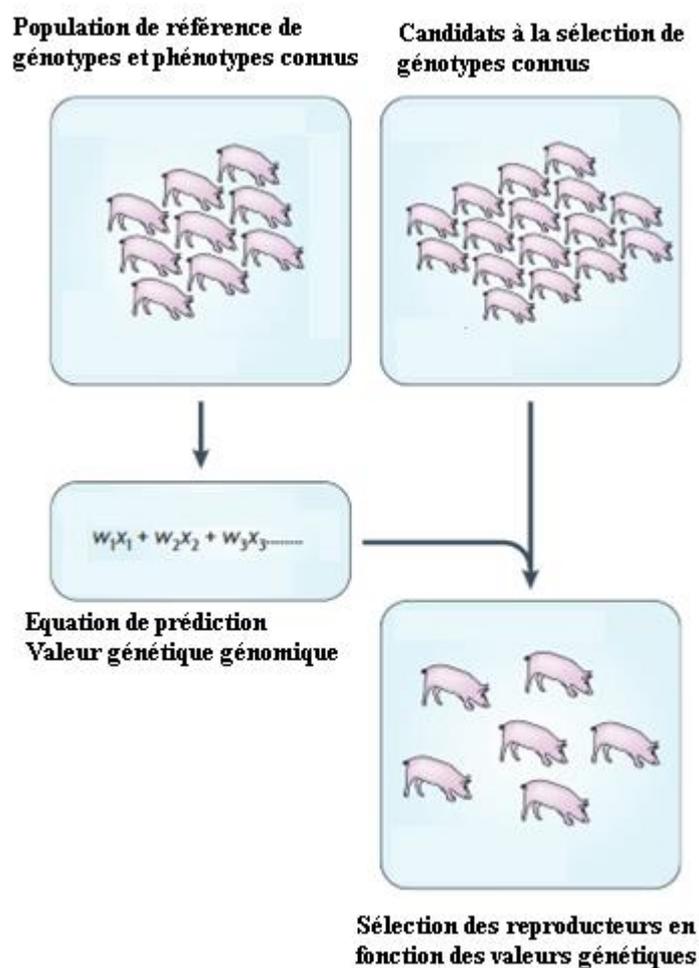


Figure I-4 : Principe de la sélection génomique (Figure adaptée de Goddard et Hayes, 2009)

4. La santé en élevage

a) Les principales maladies du porc affectant la production

Les porcs sont sensibles à de nombreux pathogènes. Ceux-ci entraînent l'apparition de maladies diverses dans les élevages qui impactent le bien-être des animaux et la productivité, et ont comme conséquences l'utilisation de médicaments (antibiotiques notamment) dont la surconsommation peut induire des résistances. Le Tableau I-3 rassemble les principales maladies affectant les porcs en élevage et ayant un effet sur le système de production. Les élevages doivent maîtriser le mieux possible tous ces risques de maladies.

b) Gestion de la santé en élevage

La santé en élevage est un aspect important du système de production porcine. Elle se caractérise par l'absence de maladie et la réalisation de performances zootechniques optimales. En premier lieu, il est capital de limiter au maximum l'apparition et la propagation de maladies au sein des élevages pour assurer le bien-être animal et la sécurité sanitaire des produits. L'apparition de maladies en élevage ayant un impact direct sur les coûts de production et l'image vis-à-vis des citoyens (scandales sanitaires, consommation d'anti-infectieux), il est essentiel de les maîtriser pour la pérennité de la filière.

L'état de santé en élevage dépend de l'équilibre entre la présence d'agents infectieux (pression d'infection) et la protection et la résistance de l'hôte (immunité, vaccination, stress...), cet équilibre étant fortement influencé par l'environnement de l'animal et de l'élevage (biosécurité, logement, alimentation...). Pour maintenir ce fragile équilibre et améliorer la santé des animaux en élevage, différents leviers d'action sont mis en place. Tout d'abord, il est important de réduire la pression en pathogènes en améliorant le niveau sanitaire et en limitant l'entrée et la propagation des agents infectieux par des contaminations externes. Cela peut être réalisé par la mise en place de règles strictes d'hygiène et la rénovation des bâtiments si nécessaire. Celles-ci ont déjà un impact positif sur le statut sanitaire des élevages mais elles nécessitent un suivi et une amélioration continue. Par exemple, l'IFIP recommande actuellement aux éleveurs de se concentrer sur la gestion des circuits véhicules- personnes- animaux, l'installation d'un sas d'entrée, l'utilisation de tenues et chaussures de travail ou encore l'augmentation de la fréquence de lavage des mains (IFIP, Institut du Porc, 2016b). Toujours selon un rapport de l'IFIP, la gestion de la qualité de l'eau doit également être optimisée, ainsi que le nettoyage et la désinfection des locaux (IFIP, Institut du Porc, 2016b).

L'accroissement de la robustesse des animaux face aux infections est identifiée comme une priorité en élevage. Le champ d'action est large et comporte l'amélioration des conditions de vie, l'apport alimentaire, l'efficacité vaccinale, la réduction du stress ou encore l'amélioration génétique par la sélection d'animaux caractérisés pour leur robustesse, leur potentiel de résistance aux infections, leurs compétences immunitaires.

Le développement de stratégies améliorant la santé est un levier important, notamment par le remplacement des traitements préventifs médicamenteux, par une détection précoce des symptômes cliniques de pathologies et un traitement plus individualisé afin d'administrer des médicaments pour le bon pathogène et avec les bonnes doses. En effet, l'utilisation d'antibiotiques doit être raisonnée afin de limiter le développement de l'antibiorésistance qui constitue désormais un enjeu de santé publique au niveau mondial. Ce phénomène a entraîné une action globalisée mise en place dès les années 2000.

Le plan pluriannuel ECOANTIBIO (Debaere, 2016) du Ministère de l'Agriculture (2012-2017) en est l'un des piliers. Son principal objectif était la réduction de 25% de l'usage des antibiotiques en médecine



vétérinaire en cinq ans. Pour cela, il encourageait à promouvoir les bonnes pratiques et à sensibiliser les différents acteurs, développer les alternatives évitant les recours aux antibiotiques, renforcer l'encadrement des pratiques et des règles de prescription commerciales, améliorer le dispositif de suivi de la consommation des antibiotiques et de l'antibiorésistance et promouvoir la même approche à l'échelon européen et international. Cette première version du plan ECOANTIBIO a été un succès avec une réduction de 37% d'exposition des animaux aux antibiotiques en cinq ans (2016 par rapport à 2011). Le deuxième plan ECOANTIBIO2 2017-2021 (<http://agriculture.gouv.fr/le-plan-ecoantibio-2-2017-2021>) doit permettre d'inscrire dans la durée ces résultats très encourageants.

Des projets associés se développent comme la démarche pour la Gestion des traitements Vétérinaires en Elevage (GVET) qui tente de moderniser le registre des traitements et d'assurer une mesure standardisée et officielle des usages d'antibiotiques (IFIP, Institut du Porc, 2016b). Un autre projet en cours vise à construire un outil capable de réaliser une détection précoce des pathologies en fonction du comportement alimentaire et de l'abreuvement des animaux qui peuvent être modifiés avant même l'apparition des symptômes pour une prise en charge plus rapide. La publication du Guide de Bonnes Pratiques d'Hygiène (GBPH) a pour objectif d'aider les éleveurs à maîtriser la sécurité sanitaire de la viande de porc (IFIP, Institut du Porc, 2016b).



Tableau I-3 : Classement et caractéristiques des principales maladies porcines

Classement	Maladie	Agent étiologique	Caractéristiques principales
Maladies réglementées	Aujeszky	<i>Suid herpesvirus 1</i>	Hôtes : suidés domestiques et sauvages, éradiquée dans les élevages en Europe Contamination : oro-nasale en élevage, vénérienne chez les suidés sauvages Localisation dans le ganglion trigéminal où il entre en phase de latence et sera réactivé par un stress Symptômes: mortalité par méningo-encéphalite chez les porcelets de -15j ; syndrome grippal pour les porcs en engraissement (mortalité 10%) + surinfections bactériennes
	Peste Porcine classique ou choléra du porc	Pestivirus de la famille des <i>Flaviviridae</i>	Hôtes : suidés domestiques et sauvages Transmission : directe ou indirecte (déchets d'origine porcine) et transmission verticale <i>in utero</i> Symptômes : mortalité brutale (forme suraiguë) ; perte d'appétit, hyperthermie, plaques hémorragiques et mortalité de plus de 80% en 5 à 10 jours (forme aiguë) ; mortalité inférieure à 50% en 30 à 60 jours, entraîne avortement ou mort-nés (forme atténuée) Vaccination disponible pour circonscrire la maladie, pas de traitement, contrôle par une prophylaxie sanitaire stricte
	Brucellose*	Bactéries du genre <i>Brucella</i>	Hôtes : bovins, porcs, ovins, caprins, équines, camélidés, chiens, hommes, ... Transmission : concentration élevée des bactéries dans les eaux fœtales et le lait Symptômes : chez les animaux, généralement bénigne mais provoque avortement et mort-nés Zoonose transmise facilement à l'homme par pénétration de la peau et des muqueuses, entraînant fièvre, céphalée, sudation, frissons, perte de poids, douleurs généralisées
	Diarrhée épidémique porcine	<i>Alphacoronavirus</i>	Symptômes : Diarrhée liquide, vomissements, anorexie, mortalité de 50% à 100% chez les porcelets, maximum 5% chez les animaux plus âgés
	Peste porcine Africaine	Virus de la famille des <i>Asfaviridae</i>	Hôtes : suidés domestiques et sauvages Transmission : très contagieuse, endémique en Afrique subsaharienne et en Sardaigne Symptômes : hémorragies souvent fatales Pas de traitement ni vaccin, importance de la prévention par des mesures d'importation rigoureuses et élimination des déchets alimentaires dans les transports
Maladies à fort impact économiques	Syndrome dysgénésique et respiratoire porcin	Virus du genre des <i>Arterivirus</i>	Maladie de l'oreille bleue qui touche les porcs domestiques Symptômes : troubles de la reproduction chez les truies et pneumonie des porcelets, accroissement de la sensibilité aux infections bactériennes secondaires Transmission : par déplacement d'animaux infectés, fèces, urine, semence, insectes, dispersion aérienne

	Maladie d'amaigrissement du porcelet	<i>Circovirus</i> porcin de type 2 (PCV2)	Hôtes : suidés (porcelets de 7 à 15 semaines) Symptômes : fièvre et perte d'appétit, difficultés respiratoires, diarrhées, parfois la mort Impact maîtrisé par des améliorations des conditions d'hygiène et réduction des situations de stress
	Pneumonie enzootique	<i>Mycoplasma hyopneumoniae</i>	Hôtes : pores à l'engraissement et porcelets sevrés Symptômes : toux, faible mortalité Transmission aérienne et le trafic des animaux, très contagieuse
	Complexe respiratoire porcin	<i>Mycoplasma hyopneumoniae</i> + surinfections	Entité clinique caractérisée par des troubles respiratoires, retard de croissance, augmentation de l'indice de consommation, augmentation chez les porcs en post-sevrage et/ou engraissement
	Pleuropneumonie porcine ou actinobacillose	<i>Actinobacillus pleuropneumoniae</i>	Hôtes : porcelets et porcs à l'engraissement Symptômes : pneumonie et pleurésies hémorragiques ; forte fièvre, mortalité en 24h (forme suraiguë) ; manque d'appétit, fièvre modérée, respiration accélérée et difficile (forme aiguë) ; sans signe clinique (forme chronique) Transmission : contact direct, toux
Maladies d'élevage	Complexe digestif	Différents agents	Syndrome de la diarrhée grise post-sevrage Multifactoriel, apparition liée aux conditions d'élevage, modes d'alimentation
	Colibacillose	<i>E. coli</i>	Trois formes : colibacillose de sevrage (diarrhée, retard de croissance), colitoxicose (mortalités sporadiques, déshydratation sévère, thrombose, coagulation intravasculaire), maladie de l'œdème (multiplication des bactéries dans l'intestin produisant une vérotoxine et entraînant des œdèmes)
Zoonoses	Grippe porcine*	<i>Influenza virus A</i>	Hautement contagieuse, morbidité de 100% rétablissement en 3 à 7 jours sans infection secondaire Symptômes : fièvre, léthargie, anorexie, perte de poids, difficultés respiratoires, toux Infection occasionnelle chez l'homme
	Trichinellose*	<i>Trichinella spp</i> (ver)	Hôtes : nombreux animaux sauvages et domestiques, hommes Infection : consommation de viande contaminée par des larves Symptômes : pas de manifestation cliniques chez le porc mais grave chez l'homme
	Salmonellose*	Bactéries du genre <i>Salmonella</i>	Hôtes : mammifères, oiseaux Symptômes : asymptomatique chez le porc mais cause de la moitié des cas de toxi-infections alimentaires chez l'homme, avec une gastro-entérite aiguë, mortelle pour les personnes au système immunitaire affaibli

* = risque zoonotique (transmission à l'homme)



B. Caractérisation de la compétence immunitaire

1. Contexte et définition

Au cours des trente dernières années, comme décrit précédemment, les systèmes de sélection en élevage porcin se sont concentrés avec succès sur les caractéristiques de production telles que le taux de croissance, la composition de la carcasse, l'efficacité alimentaire et la qualité de la viande, ainsi que la prolificité (Tribout *et al.*, 2013). Pendant ce temps, le développement de cartes génétiques a permis l'identification de nombreux QTL (pour *Quantitative Trait Loci* en anglais) dans le génome du porc (Bidanel *et al.*, 2000 ; <https://www.animalgenome.org/cgi-bin/QTldb/SS/index>). La sélection génétique et l'amélioration de l'environnement ont été concomitantes de l'intensification et la systématisation de l'insémination artificielle. Dans les populations porcines actuelles très sélectionnées, certaines maladies sont évitées par la vaccination, l'antibiothérapie et le respect de règles sanitaires strictes. Cependant, au cours de cette période de forte augmentation de la valeur génétique des porcs, de nouvelles maladies sont apparues, entraînant d'importantes pertes économiques telles que le syndrome dysgénésique et respiratoire porcin (SDRP) et le syndrome de dépérissement multisystémique post-sevrage.

Les animaux de production sont élevés dans des fermes conventionnelles et, en dépit de règles sanitaires strictes, ils sont confrontés à de nombreux agents pathogènes présents dans l'environnement. Il existe plusieurs options complémentaires pour lutter contre les infections : approches ciblées par pathogène (Kadowaki *et al.*, 2012; Shimazu *et al.*, 2014) ou approches globales visant à caractériser la compétence immunitaire individuelle, en tant qu'information de base potentiellement prédictive de la capacité de réponse aux pathogènes et à la vaccination (Calder, 2007; Visscher *et al.*, 2002; Wilkie et Mallard, 1999). Caractériser l'immunocompétence individuelle est donc une des pistes de recherche mise en place pour identifier des leviers d'action qui influent sur la santé des animaux en élevage. Dans ce contexte, le phénotypage des paramètres immuns et l'analyse de la part génétique des variations observées est un pré-requis nécessaire pour évaluer la faisabilité d'inclure des caractères d'immunité dans les schémas de sélection.

L'étude de la variabilité des capacités immunitaires chez des individus sains est rarement faite à grande échelle en production porcine. C'est pourtant une étape essentielle pour identifier des informations prédictives robustes et étudier à quel point l'hétérogénéité dans la réponse pathogénique ou pharmacogénique est liée aux capacités immunitaires individuelles pré-existantes. Des projets de grande envergure ont été lancés chez l'homme sur cette thématique, comme par exemple le projet ImmVar (De Jager *et al.*, 2015) ou le projet Milieu Intérieur (Thomas *et al.*, 2015), qui analyse la variabilité des paramètres immunitaires d'une cohorte de 1000 individus sains (sans signes cliniques d'infection). Les projets ANR



IMMOPIG et SUS_FLORA sur lesquels s'appuie le projet de thèse s'inscrivent dans cette démarche (voir Annexe 1).

a) Aperçu de la complexité du système immunitaire

Cette section a pour principal but de présenter les bases biologiques qui conduisent aux mesures réalisées pour qualifier la réponse immunitaire constituant l'immunophénotypage. Le système immunitaire est l'ensemble des organes, tissus, cellules et molécules assurant l'immunité d'un organisme (source : définition Larousse). Il fonctionne grâce à un ensemble de cellules hétérogènes dans leurs formes, leurs fonctions, leur capacité à se multiplier et à se transformer, ainsi que dans la durée de leur vie (Albers *et al.*, 2005; Calder, 2005; Calder and Kew, 2002). Le système immunitaire permet aux organismes de se protéger contre les pathogènes et de maintenir leur intégrité par reconnaissance du soi et du non soi. Ainsi, une réponse immunitaire mal dirigée ou mal maîtrisée peut conduire à des maladies auto-immunes, des allergies, favoriser des cancers.

Face à un pathogène, la réponse immunitaire est classiquement subdivisée en deux temps : la réponse innée non spécifique et la réponse adaptative qui apporte une mémoire immunologique lors d'une exposition ultérieure au même pathogène. Les épithéliums externes (ex. : peau) et internes (ex. : muqueuse intestinale) jouent un rôle de protection en agissant comme des barrières physiques et fonctionnelles. La réponse innée, immédiate, est relayée après quelques heures voire quelques jours, par une réponse adaptative liée à l'activation de cellules de l'immunité dédiées (Figure I- 5).

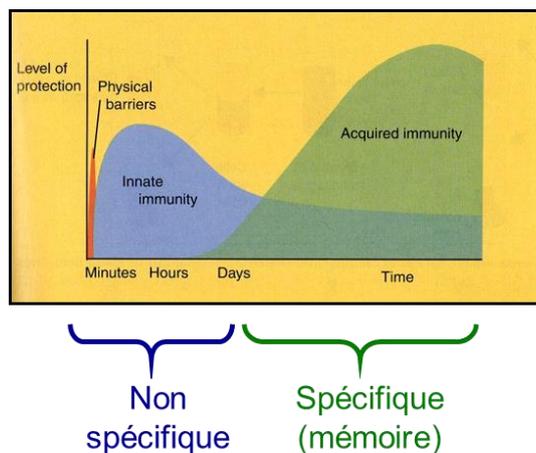


Figure I-5 : Organisation temporelle des réponses immunitaires innées et adaptatives (source : Delves P.J., Roitt's Essential Immunology)



b) Organisation

Le système immunitaire comprend le sang et la lymphe qui sont circulants, et les organes lymphoïdes primaires et secondaires (Figure I-6, Tableau I-4).

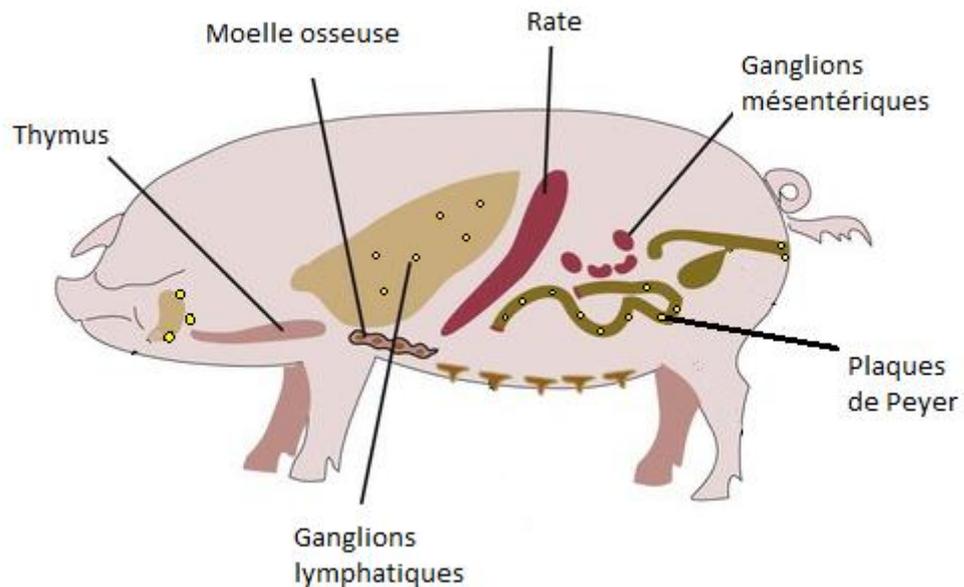


Figure I-6 : Organes lymphoïdes du porc (Figure adaptée de Roitt's Essential Immunology et veterinariankey.com)

Les organes lymphoïdes primaires sont le thymus et la moelle osseuse, où les lymphocytes T et B se différencient à partir des progéniteurs lymphoïdes, prolifèrent et sont sélectionnés. Les cellules progénitrices sont, elles-mêmes dérivées des cellules souches hématopoïétiques (Figure I-7). Les cellules progénitrices des lymphocytes T et B sont produites dans la moelle osseuse.

Tableau I-4 : Organisation et fonctionnalités des organes lymphoïdes primaires et secondaires

Groupes	Dénomination	Localisation	Fonctionnalités
Organes lymphoïdes primaires	<i>Moelle osseuse</i>	Partie centrale des os courts et plats	Activité hématopoïétique (production des différentes lignées de cellules sanguines)
	<i>thymus</i>	Partie antéro-supérieure de la cage thoracique	Rôle primordial dans la différenciation des lymphocytes T ; contient des cellules dendritiques, des thymocytes (cellules lymphoïdes immatures provenant de la moelle), des cellules épithéliales, des macrophages : régresse chez l'adulte
Organes lymphoïdes secondaires	<i>Ganglions lymphatiques</i>	Répartis dans tout l'organisme, groupés en aires ganglionnaires	Lieu de prolifération et de différenciation des lymphocytes B et T, filtre de la circulation lymphatique. Structure inversée chez le porc, les follicules lymphoïdes se distribuent au centre du ganglion, la médulla occupe une position périphérique La lymphe entre via les vaisseaux lymphatiques afférents qui pénètrent depuis la capsule dans le cortex alors qu'elle sort par les vaisseaux lymphatiques efférents. Les lymphocytes T entrent et retournent dans la circulation systémique principalement à travers les veinules endothéliales hautes.
	<i>Rate</i>	Organe abdominal intra-péritonéal	Pulpe rouge : rôle important dans la régulation de la formation et de la destruction des éléments figurés du sang, notamment des hématies. Pulpe blanche : lieu des rencontres antigènes-lymphocytes (lymphocytes T dans la gaine lymphoïde péri-artérielle et lymphocytes B dans le corpus de Malpighi)
	<i>Amygdales</i>	Situées dans la gorge à l'entrée des voies respiratoires sur le pourtour du pharynx	Composées de follicules lymphoïdes contenant des lymphocytes B et T entre les follicules.
	<i>Plaques de Peyer</i>	Au niveau de la paroi intestinale dans la partie terminale de l'intestin grêle	Présence de lymphocytes B et de lymphocytes T plus diffuse Partie centrale du dôme contenant des cellules M formant une cavité intraépithéliale où se trouvent des macrophages, des cellules présentatrices d'antigènes et des lymphocytes.



Tout le développement et la maturation des lymphocytes B a lieu dans la moelle osseuse, les lymphocytes T immatures (précurseurs T) quittent la moelle osseuse et terminent leur maturation dans le thymus. La génération des lymphocytes T et B dans les organes lymphoïdes primaires est suivie de leur migration dans les organes lymphoïdes secondaires via le sang, où ils pourront être activés s'ils rencontrent les antigènes dont ils sont spécifiques. S'ils ne rencontrent pas l'antigène dont ils sont spécifiques, les lymphocytes naïfs (c'est-à-dire n'ayant pas été activés) continueront de circuler.

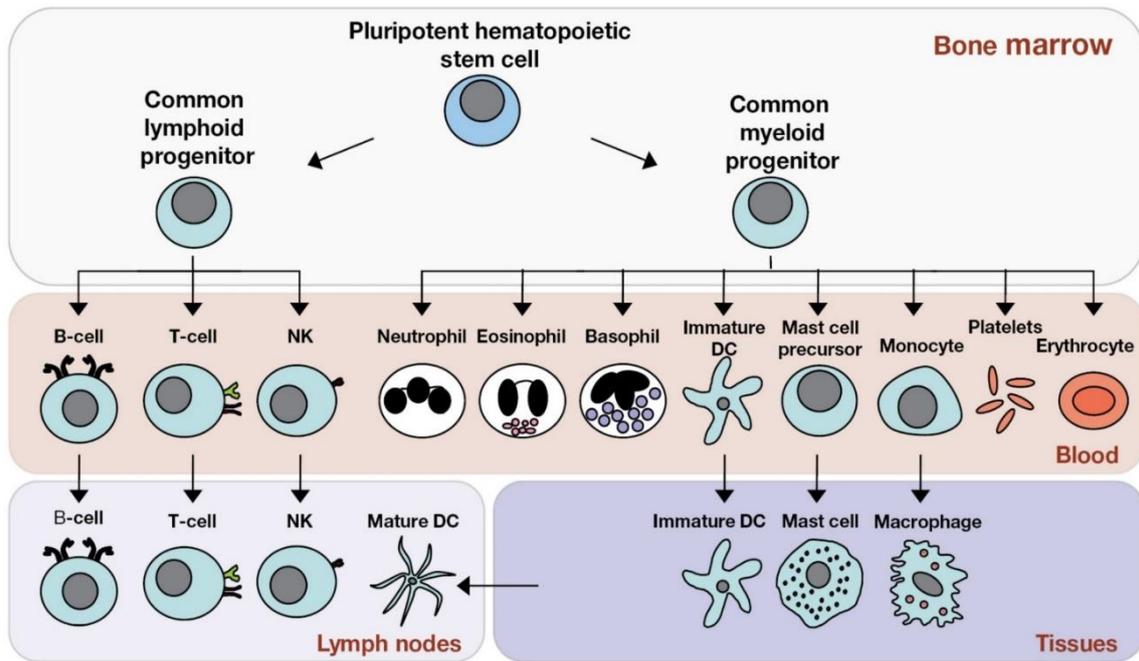


Figure I-7 : Origine et différenciation des cellules immunitaires (source : Roitt's Essential Immunology)

Les organes lymphoïdes secondaires sont des lieux de concentration des lymphocytes, au niveau desquels s'effectue l'activation de la réponse immunitaire adaptative, autrement dit l'activation des lymphocytes naïfs qui se différencieront en cellules effectrices et cellules mémoires. On distingue des organes lymphoïdes secondaires bien structurés, comme la rate et les ganglions lymphatiques, et des tissus lymphoïdes, accumulations de cellules lymphoïdes non encapsulées dans un organe, comme les tissus lymphoïdes associés aux muqueuses (MALT pour « *Mucosa-Associated Lymphoid Tissue* ») comprenant les amygdales et les plaques de Peyer. Plus précisément les plaques de Peyer sont associées au GALT (Gut-Associated Lymphoid Tissue). L'organisation des plaques de Peyer est présentée à la Figure I-8.



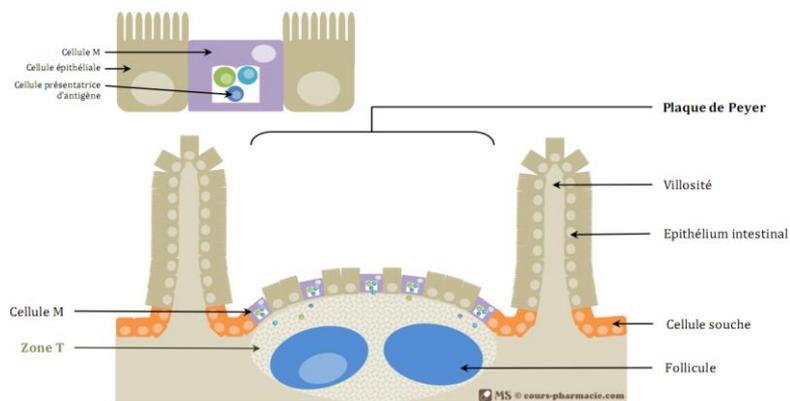


Figure I-8 : Schéma simplifié représentant la structure tissulaire d'une plaque de Peyer (source : Roitt's Essential Immunology)

c) Système immunitaire inné

Le Tableau 4 récapitule les principales cellules de l'immunité innée et leurs fonctions. Les effecteurs moléculaires majeurs de la réponse innée sont le $TNF\alpha$, les interleukines (IL) IL-1, IL-6, IL-8, IL-23, IL-17. L'immunité innée est également médiée par les anticorps naturels. Ce sont des immunoglobulines à épitopes homologues très conservés, qui ne nécessitent pas de stimulation extérieure du système immunitaire pour être sécrétés en grande quantité. Ils constituent la première barrière efficace contre l'invasion des pathogènes notamment par leur présence abondante dans l'organisme de l'hôte et leur poly-réactivité (Siwek *et al.*, 2015).

Tableau I-5 : Principales cellules de l'immunité innée

Cellules de l'immunité innée	Fonctions
Monocytes	Cellules sanguines immatures qui se différencient en macrophages ou cellules dendritiques dans les tissus
Macrophages	Dans les tissus, elles phagocytent et peuvent présenter les antigènes
Cellules dendritiques	Cellules phagocytaires et cellules présentatrices d'antigènes professionnelles
Granulocytes (neutrophiles, basophiles, éosinophiles)	Les neutrophiles sont les plus nombreux dans le sang, ils sont attirés sur les sites inflammatoires et jouent un rôle dans la phagocytose, les basophiles jouent un rôle dans l'allergie et les éosinophiles ont une action antiparasitaire
Mastocytes	Dans les muqueuses, les mastocytes jouent un rôle dans l'allergie
Cellules NK (<i>Natural Killer</i>) ou cellules tueuses naturelles	Elles éliminent les cellules dont la fonction est altérée (cellules tumorales, infectées)
Lymphocytes T $\gamma\delta$	Lymphocytes avec des caractéristiques de cellules de l'immunité innée et de l'immunité adaptative
Cellules résidentes	Elles peuvent exprimer des molécules du CMH I et sécréter des médiateurs de l'inflammation



La réaction inflammatoire est l'une des premières réponses contre une infection ou une irritation. Elle est induite par des cellules immunitaires (macrophages, cellules dendritiques, mastocytes) mais également des cellules de tissus cibles de l'infection, qui expriment à leur surface des récepteurs de reconnaissance de motifs moléculaires (*Pathogen Recognition Receptor* ou PRR) associés aux pathogènes (*Pathogen-Associated Molecular Pattern* ou PAMP) ou à des débris cellulaires (*Damage-Associated Molecular Pattern* ou DAMP). Les récepteurs Toll Like (TLR) sont des PPR intensivement étudiés. Au nombre de 10 à 15 selon les espèces (13 chez l'homme, 10 chez le porc), ils ont la particularité de reconnaître des classes différentes de motifs.

d) Système immunitaire adaptatif

Le système immunitaire adaptatif a pour acteurs principaux les lymphocytes T, qui contribuent à l'immunité à médiation cellulaire, et les lymphocytes B qui sont responsables de l'immunité à médiation humorale. Ces deux populations cellulaires sont présentes majoritairement dans le sang mais également dans la lymphe, les organes lymphoïdes et les tissus.

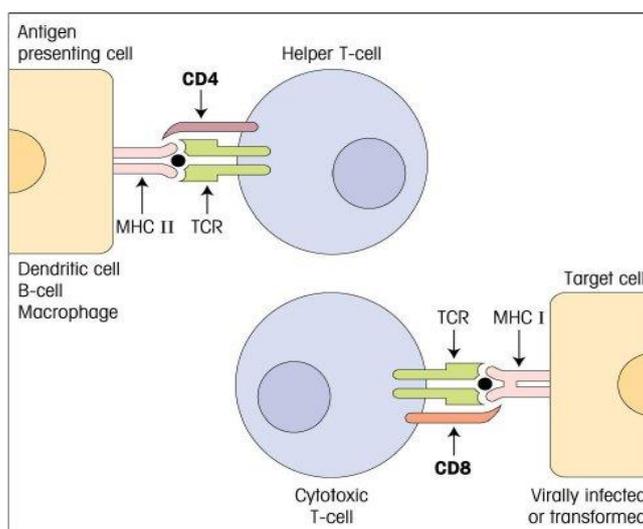


Figure I-9 : Présentation des peptides par les molécules du complexe majeur d'histocompatibilité (CMH) qui assurent la reconnaissance du soi et du non soi (Delves *et al.*, 2011)

Les antigènes sont présentés aux lymphocytes par les molécules du complexe majeur d'histocompatibilité (CMH) qui assurent la reconnaissance du soi et du non soi (Figure I-9). Les molécules de classe I présentent des peptides endogènes et viraux aux lymphocytes T cytotoxiques (CD8+) et les molécules de classe II présentent des peptides exogènes (ex : peptides bactériens) aux lymphocytes T CD4+.



Les lymphocytes T cytotoxiques induisent la mort par apoptose des cellules infectées par des virus (ou par d'autres agents infectieux) ou des cellules cancéreuses après activation par la reconnaissance d'un complexe peptide-CMH I présenté à la surface d'une cellule cible. Cette activation entraîne leur prolifération ou expansion et leur dissémination à la recherche d'une cible. Ils relarguent des perforines qui induisent la lyse des cellules cibles par la formation de pores dans leur membrane et des granzymes qui entrent dans les cellules pour déclencher leur apoptose. Puis, la plupart des lymphocytes T cytotoxiques meurent par apoptose et sont éliminés par les phagocytes mais un petit nombre de lymphocytes deviennent des lymphocytes T cytotoxiques mémoire.

Les lymphocytes T auxiliaires, ou lymphocytes T CD4+, orchestrent la réponse immunitaire. Ils reconnaissent les peptides présentés par le CMH II sur les cellules présentatrices d'antigènes et relarguent des cytokines pour influencer l'activité de nombreux types cellulaires et notamment activer les lymphocytes T cytotoxiques ou les lymphocytes B. Une partie des lymphocytes T auxiliaires est conservée pour constituer une population de lymphocytes T mémoire.

Les lymphocytes B produisent les anticorps qui circulent dans le sang et la lymphe pour l'immunité à médiation humorale. Ils peuvent reconnaître un antigène sans cellule présentatrice et sont activés par un lymphocyte T auxiliaire pour se différencier en plasmocytes. Les plasmocytes sont des cellules à courte durée de vie qui sécrètent des anticorps. Près de 10 % des plasmocytes survivent à l'issue d'une infection et deviennent des lymphocytes B mémoire à longue durée de vie. Les anticorps, ou immunoglobulines (Ig), sont des protéines permettant la reconnaissance et la neutralisation de corps étrangers comme les agents infectieux ou les toxines. Ces anticorps peuvent se fixer aux microorganismes, et fournissent ainsi des repères aux cellules phagocytes et au système du complément.

La Figure I-10 illustre l'activation de la cascade immunitaire suite à une infection par un pathogène, les cellules dendritiques assurant un rôle d'activation de la réponse adaptative par présentation des antigènes aux cellules naïves (Th0) et induisant leur différenciation en voie Th1 à médiation cellulaire ou Th2 à médiation humorale (production d'anticorps). Le système du complément participe à la défense de l'organisme par une cascade biochimique du système immunitaire qui facilite l'élimination des pathogènes par d'autres cellules du système immunitaire.



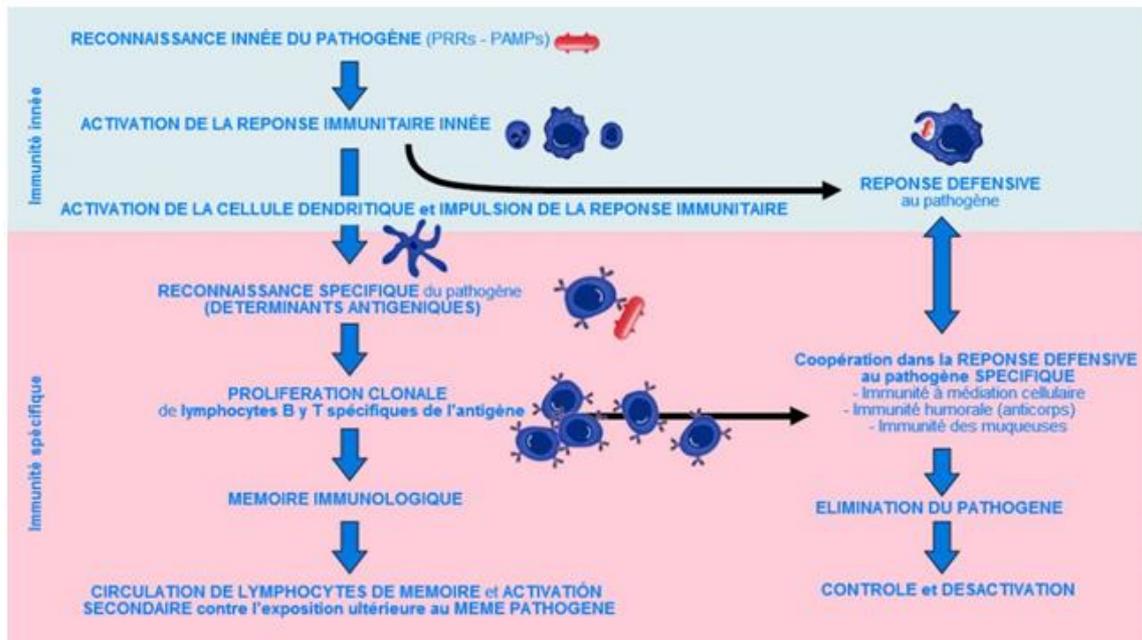


Figure I-10 : Mise en place de la réaction immunitaire (source : Le site de la filière porcine, 3trois3.com)

e) Phénotypage des paramètres immunitaires

Qualifier la capacité immunitaire requiert de phénotyper des paramètres immunitaires. Le système immunitaire est à la fois complexe et multi-acteurs, ce qui implique de déployer une large palette de mesures pour avoir une information globale de l'immunocompétence d'un individu (Albers *et al.*, 2005; Calder, 2005; Calder et Kew, 2002). Pour analyser des paramètres immunitaires, il est possible d'envisager différentes approches expérimentales : (i) travailler *in vivo* ou *in vitro* ; (ii) comparer des individus challengés/non challengés avec un pathogène ou vaccinés/non vaccinés ; (iii) comparer des tissus traités/non traités d'un même individu (Schroyen et Tuggle, 2015). Il faut pouvoir caractériser la capacité immunitaire à l'état basal mais également en réponse à des pathogènes et des facteurs abiotiques environnementaux (Chaussabel *et al.*, 2010).

f) Le sang : un tissu privilégié pour l'immunophénotypage

Le sang correspond à environ 8 % du poids du corps (Fan et Hegde, 2005) et contient de multiples types cellulaires à différents stades de leur cycle de vie. Chez le porc, 98% des éléments constitutifs du sang sont des érythrocytes (globules rouges), 1% des plaquettes, et moins de 1% des leucocytes (Choi *et al.*, 2014). Ils baignent dans le plasma qui contient des facteurs coagulants (fibrines) et le sérum. On trouve également dans le sang de nombreuses microvésicules et molécules bio-actives (oxygène, nutriments, métabolites,



anticorps, cytokines, hormones), ainsi que des molécules médicamenteuses et toxines exogènes (Mohr et Liew, 2007).

Le sang circulant conduit les cellules immunitaires naïves ou éduquées d'un site à un autre. Comme l'illustre la Figure I-11, le sang est la voie de circulation des cellules immunitaires dans l'organisme pour notamment atteindre les ganglions lymphatiques où les réponses immunitaires spécifiques se mettent en place et rejoindre ensuite les tissus pour agir au niveau des sites de l'inflammation (Chaussabel *et al.*, 2010). Les cellules circulantes dans le sang peuvent être utilisées comme des sentinelles répondant aux changements macro et micro environnementaux des organes. Le sang est considéré comme un tissu privilégié pour analyser le statut immunitaire individuel (Liew *et al.*, 2006) et est largement utilisé comme tel.

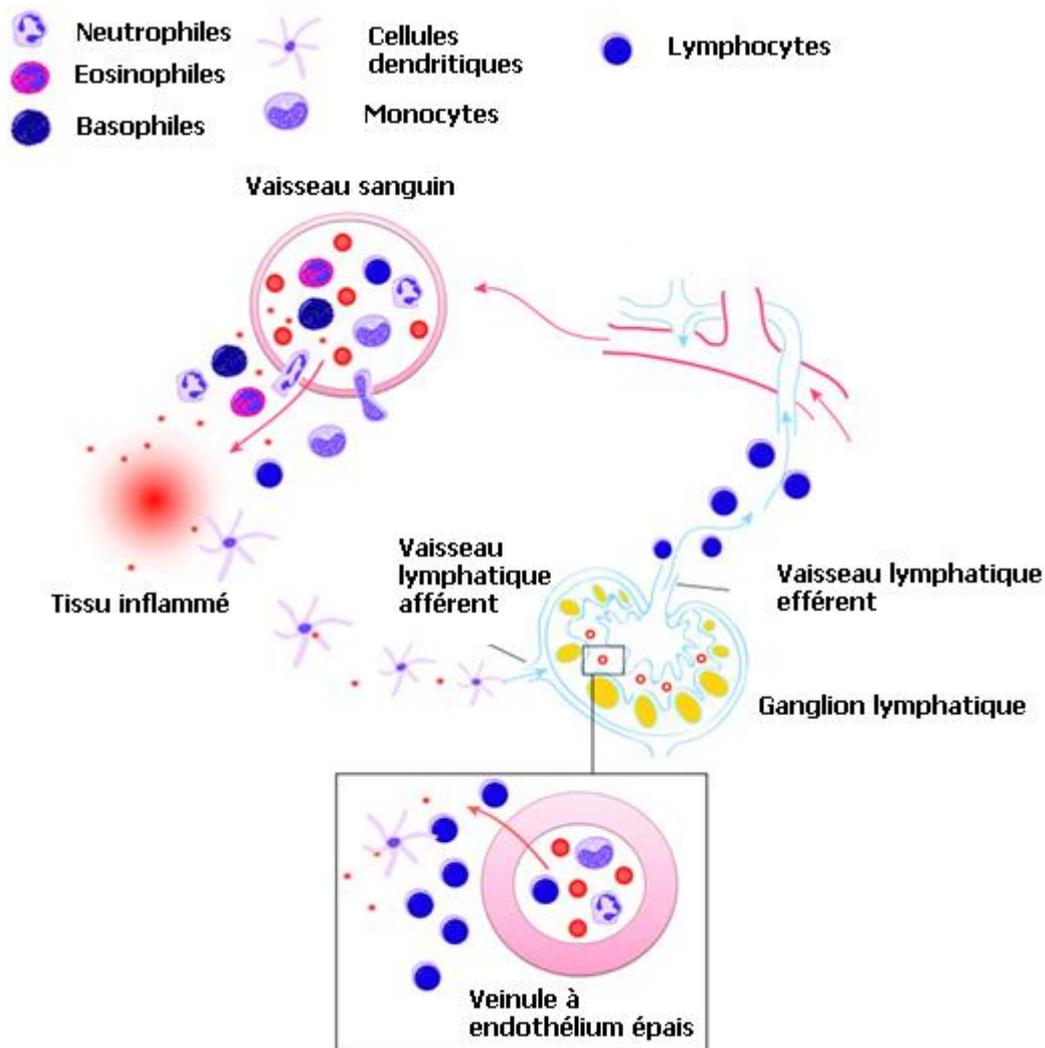


Figure I-11 : Représentation de l'implication du sang dans le système immunitaire (Figure traduite de Chaussabel *et al.*, 2010)



Le sang circulant est facile à prélever, et permet un échantillonnage peu invasif et peu risqué pour l'animal. Il favorise des prélèvements répétés sur un individu pour le suivi d'une infection ou d'un traitement.

g) Dosages sanguins et prélèvements associés

Il est possible d'étudier le sang dans sa totalité ou par fractions, en adaptant les prélèvements et les méthodes de préparation. En fonction de l'objet de l'étude, on peut se concentrer sur le plasma, le sérum (plasma sans les facteurs coagulants) ou les populations cellulaires. La Figure I-12 représente les différents types de mesures qui peuvent être effectués à partir d'échantillons de sang.

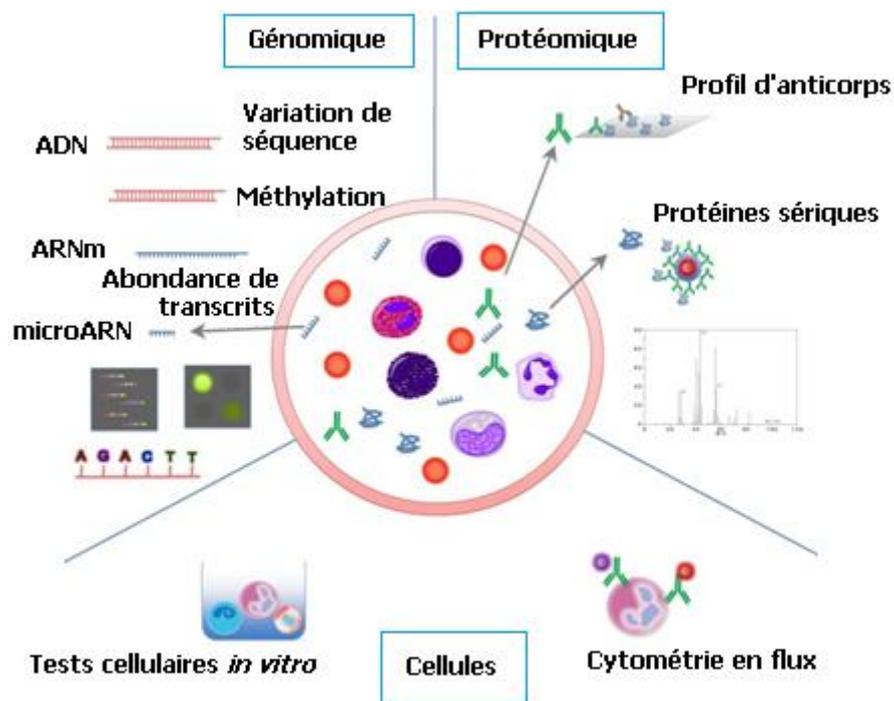


Figure I-12 : Ensemble des techniques applicables sur le sang pour phénotyper l'immunité (Figure adaptée de Chaussabel *et al.*, 2010)

Les traits immunitaires mesurés à partir du sang peuvent être répartis en deux grandes catégories :

- les traits immunitaires descriptifs de l'état de l'individu au moment du prélèvement (ex. : numération-formule sanguine (NFS), comptage de sous-populations cellulaires par cytométrie de flux, dosage de protéines sériques et des taux d'anticorps) ;



- les traits immunitaires de réponse, mesurés après stimulation *in vitro* du sang ou de populations cellulaires (ex. : phagocytose, dosage de cytokines dans les surnageants de culture, mesure de prolifération cellulaire).

Chaque étude utilise un portefeuille différent de traits immunitaires selon les objectifs et les réactifs disponibles pour l'espèce considérée. Par exemple, Duffy *et al.* (2014), pour qualifier la variabilité individuelle de la compétence immunitaire chez des hommes sains (non malades), ont mis au point des tests de stimulation *in vitro* du sang total avec 27 stimuli différents, dont des bactéries gram négatives, gram positives, des mycobactéries, des champignons et virus vivants et ont mesuré les taux de réponses par dosage de différentes cytokines (IFN- γ , IFN- β , TNF- α , IL-1 β , IL-23). Ils ont également développé des systèmes pour étudier la réponse à des ligands purifiés ou des ligands synthétiques pour les voies de réponses innées dont les TLR, NLR (nucleotide-binding domain and leucine-rich repeat containing molecules), RLR (RIG-I-like receptors) et CLR (C-type lectin like receptors).

Chez le porc, des réactifs et appareils de mesure sont disponibles pour doser à grande échelle différents paramètres immunitaires (Flori *et al.*, 2011 et le chapitre 3 des résultats). Le Tableau présenté en Annexe 2 résume les principaux caractères immunitaires mesurables chez le porc, les méthodes de mesures et les interprétations qui en sont faites.

h) Transcriptome du sang

Le transcriptome du sang est un phénotype intermédiaire largement exploré pour identifier des profils d'expression à relier à des variations physiologiques, notamment immunitaires, ou à des pathologies. Il est en effet corrélé à celui de nombreux tissus. Mohr et Liew (2007) ont ainsi observé que chez l'homme, les cellules du sang partagent 80 % de leur transcriptome avec huit autres tissus : cerveau, cœur, rein, foie, poumon, prostate, rate, estomac. Liew *et al.* (2006) ont détecté l'expression de gènes dans le sang que l'on pensait spécifiques de tissus non sanguins. Ainsi l'étude du transcriptome du sang chez l'homme a permis d'identifier des profils d'expression associés à des phénotypes de diverses natures : maladies cardiovasculaires, hypertension, arthrose juvénile, cancer, fatigue chronique, troubles neurologiques, lupus, transplantation, troubles schizophréniques et bipolaires (Liew *et al.*, 2006).

Le transcriptome du sang peut être analysé de façon globale par utilisation de puces d'expression ou par séquençage d'ARN. Ces approches permettent d'étudier les gènes qui codent des protéines mais également les ARN non codants dont les microARN. Les réticulocytes qui contiennent de l'ARN messager peuvent



contribuer jusqu'à 70 % de la quantité d'ARN du sang, notamment par des ARN messagers de globine (Fan et Hegde, 2005). Cette grande abondance peut affecter la capacité à détecter les autres espèces de transcrits par RNA-Seq, en particulier ceux avec une faible expression. Choi *et al.* (2014) ont proposé un protocole de déplétion de la globine chez le porc pour des analyses de séquençage d'ARN augmentant la sensibilité et la reproductibilité des analyses. Il est également possible de pallier cette surabondance de transcrits de globine en réalisant un séquençage à une plus grande profondeur.

Le choix du tube de collecte influence beaucoup les profils d'expression du sang périphérique. Les deux systèmes commerciaux les plus souvent utilisés pour la collecte du sang en tube et sa stabilisation sont PAXgene Blood RNA Systems (PreAnalytiX) et Tempus Blood RNA tubes (Applied Biosystems ; Menke *et al.*, 2012). Le système PAXgene contient une solution qui inhibe la dégradation de l'ARN quand le sang est introduit dans le tube. Il est difficile de superposer les profils d'expression correspondant à des ARN collectés avec deux types de tubes différents. Il faut donc utiliser le même type de tube tout au long d'un même projet pour éviter un biais de détection (Menke *et al.*, 2012).

Une autre difficulté pour analyser le transcriptome du sang est liée au fait que le sang est parmi les tissus les plus variables de l'organisme. Ainsi, la variation d'expression observée dans le sang peut refléter à la fois la variabilité entre individus et la variabilité intrinsèque à l'individu. La variation de composition du sang est principalement liée aux changements de proportions des différents composants (cellulaires, moléculaires). Cela contribue dans une proportion significative à la variation du transcriptome (Fan et Hegde, 2005). La minimalisation de la variation des échantillons est donc essentielle (Shin *et al.*, 2014). Pour cela, il faut bien maîtriser le design de l'expérience, en augmentant notamment le nombre d'échantillons, en utilisant des critères stricts d'analyse et en standardisant les prélèvements (Fan et Hegde, 2005). Whitney *et al.* (2003) ont réalisé une étude des profils d'expression des gènes dans le sang pour des individus sains. Ils ont observé une constance de ces profils mais aussi mis en évidence des profils d'expression distincts pour chaque individu, qui varient dans le temps. Afin de limiter la variabilité d'origine technique, le projet ImmVar (De Jager *et al.*, 2015) préconise de faire plusieurs prélèvements, espacés dans le temps d'un jour ou deux afin de pouvoir prendre en compte les effets de variance liés à la technique ou à l'individu. Il faut également faire des prélèvements séparés par de longs intervalles (plusieurs semaines ou mois) permettant de mesurer la variabilité environnementale (De Jager *et al.*, 2015).

Des méthodes ont été développées pour séparer la variabilité d'expression des transcrits de la variabilité de composition du sang. On peut par exemple isoler les profils transcriptomiques de chaque sous-population



cellulaire, permettant ainsi d'identifier les transcrits faiblement exprimés dans ces cellules (Palmer *et al.*, 2006). Par exemple, le projet ImmVar a permis d'étudier l'impact de la variation génétique commune sur le transcriptome de cellules lymphoïdes et myéloïdes naïves *ex vivo* et *in vivo*, en réponse à différents stimuli (De Jager *et al.*, 2015). Les gènes composant les profils d'expression cellule spécifique peuvent servir de biomarqueurs pour estimer les abondances relatives des cellules. Un des enjeux actuels est la prise en compte de la composition en sous-populations cellulaire du sang, sur la base des profils d'expression dans le sang total selon le principe de la déconvolution (Urrutia *et al.*, 2016).

Chez le porc, le premier atlas d'expression du génome a été publié en 2012 (Freeman *et al.*, 2012) et peu de temps après c'est l'immunome porcin qui a été bien caractérisé grâce à un effort international d'annotation de plus de 1400 gènes, ce qui apporte des informations utiles pour l'interprétation des données de transcriptome sanguin (Dawson *et al.*, 2013). Plus récemment, c'est le transcriptome du sang qui a été décrit par RNA-Seq (Liu *et al.*, 2017). Comme mis en évidence par D. Chaussabel (2015), les études du transcriptome du sang pour la recherche en immunologie se sont multipliées ces dernières années chez l'Homme mais également chez les animaux domestiques et notamment le porc. Ainsi, Mach *et al.* (2013) ont mis en évidence que le transcriptome sanguin est informatif pour qualifier quatre traits immunitaires : taux de cellules CD4-CD8+, capacité phagocytaire mesurée par test *in vitro*, production d'IL-2 ou IL-10 après stimulation *in vitro* du sang total. Liu *et al.* (2016) ont détecté des différentiels d'expression du transcriptome entre animaux issus de lignées divergentes pour l'efficacité alimentaire. Ces études mettent en évidence que, chez le porc comme chez l'homme, le transcriptome du sang est une source d'information potentiellement exploitable pour identifier des biomarqueurs de phénotypes variés.

2. Architecture génétique des paramètres immunitaires

De nombreuses études montrent que la capacité immune individuelle est influencée par la génétique de l'hôte, tant chez l'homme (Cho, 2015; Nédélec *et al.*, 2016; Orrù *et al.*, 2013; Quach *et al.*, 2016) que chez les animaux d'élevage, sans toutefois réduire la complexité des facteurs qui influent sur l'immunité (Brodin *et al.*, 2015).

Chez le porc, des études génétiques diverses ont révélé des variations génétiques entre races ou lignées et ont permis d'identifier des QTL et d'estimer des héritabilités pour des paramètres immuns. Il a par exemple été montré que les porcs Meishan et Large White diffèrent par plusieurs sous-populations innées de globules blancs, la capacité phagocytaire des neutrophiles et les niveaux de protéines sériques de la phase



inflammatoire aiguë (Clapperton *et al.*, 2005). Des QTL de résistance / sensibilité au virus de la pseudorage ont également été rapportés entre ces deux races (Reiner *et al.*, 2002).

Une vaste étude de cartographie QTL de la capacité immunitaire générale chez le porc a été menée par Edfors-Lilja et ses collègues (1998) qui ont mesuré de nombreux paramètres de la réponse immunitaire et ont confirmé des QTL associés au nombre et à la fonction des leucocytes, au taux d'hémoglobine sur les chromosomes 1 et 8 dans des croisements. Une étude ciblant la réponse immunitaire innée et la production d'anticorps après vaccination contre *Mycoplasma hyopneumoniae*, le virus de la maladie d'Aujeszky ou le syndrome dysgénésique et respiratoire du porc (SDRP) a révélé entre 21 et 46 QTL selon les niveaux de significativité dans une population F2 (Wimmers *et al.*, 2009).

Le contrôle génétique des caractères immunitaires a également été démontré par la sélection de porcs sur huit générations successives. Cette étude a permis d'estimer une héritabilité entre 0,07 et 0,27, selon les paramètres immunitaires (Wilkie et Mallard, 1999). Une héritabilité entre 0,18 et 0,62 pour les globules blancs totaux et plusieurs sous-populations de PBMC a été rapportée chez des porcs provenant de six fermes différentes (Clapperton *et al.*, 2008, 2009). Le projet IMMOPIG a permis d'estimer l'héritabilité de plusieurs dizaines de paramètres immuns sur des porcs Large White âgés de 60 jours (Flori *et al.*, 2011). Les résultats ont mis en évidence que de nombreux paramètres de réponse innée comme adaptive sont sous contrôle génétique (Figure I-13), avec des héritabilités moyennes à élevées.



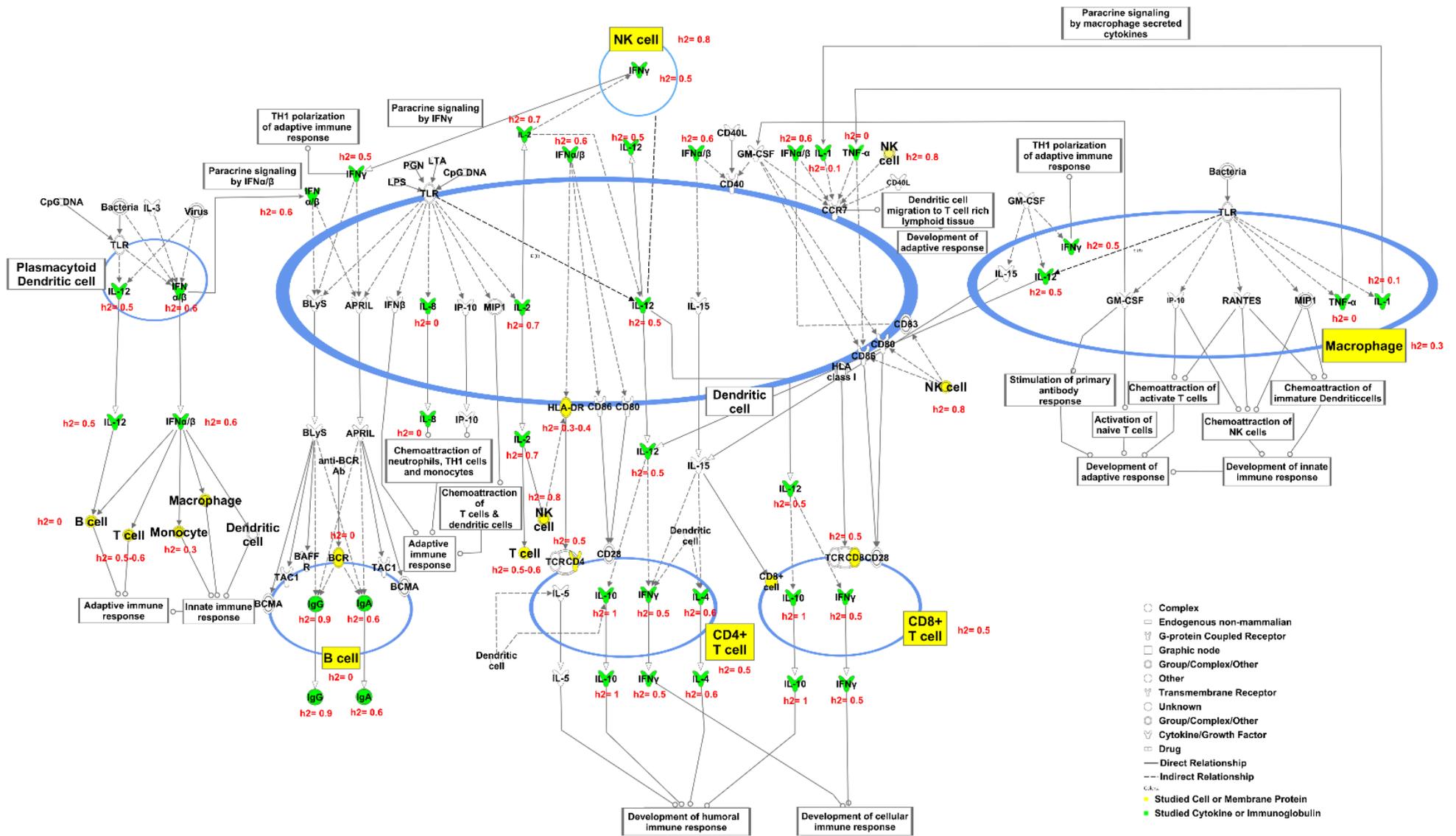


Figure I-13 : Caractères immunitaires sanguins mesurés par Flori *et al.* (2011) et héritabilités estimées

C. Cartographie du contrôle génétique de phénotypes

La recherche en génétique animale dans laquelle s'inscrit ce projet de thèse a pour but d'identifier des marqueurs génétiques associés à des traits ou phénotypes importants pour la production, qui doivent être évalués pour être inclus dans les schémas de sélection des reproducteurs. Certains de ces traits sont d'ores et déjà sélectionnés après mesures des performances des animaux (poids, épaisseur du lard dorsal, nombre de tétines...) alors que d'autres traits seront plus faciles à prendre en compte par sélection génomique et génotypage des marqueurs génétiques associés, sans mesure directe sur l'animal (efficacité alimentaire, résistance aux maladies, robustesse, comportement).

La sélection assistée par marqueurs a pu se développer grâce à l'apparition de technologies pour l'étude de marqueurs génétiques comme les microsatellites (séquence d'ADN formée par une répétition continue de motifs de 1 à 4 nucléotides), et plus tard les SNP (Single Nucleotide Polymorphisme ou polymorphisme d'un seul nucléotide). L'essor des puces de génotypage à haute densité de SNP a permis le démarrage de la sélection génomique, selon une théorie proposée par Meuwissen et Goddard dès 2001 (Meuwissen et Goddard, 2001). Dans l'approche la plus simple de la sélection génomique, une équation linéaire décrivant les effets additifs des différents marqueurs génétiques est établie pour prédire la future valeur phénotypique des animaux et donc de les sélectionner à partir de leur génotype. La méthode la plus utilisée en prédiction d'équation est GBLUP (*Genomic SNP-best Linear Unbiased Prediction*) qui suppose que les effets des SNP sont normalement distribués et ont des effets faibles (Hayes *et al.*, 2013).

La précision et la finesse des études génétiques dépendent du nombre de marqueurs utilisés et de la proportion de la variance génétique qu'ils captent. Par exemple, la valeur génétique génomique basée sur 50 000 marqueurs d'une puce de SNP 50K, entraîne une prédiction des effets des mutations causales en déséquilibre de liaison sur de larges segments chromosomiques. Mais l'accumulation de générations au sein d'une population augmente le taux de recombinaison et peut diminuer la précision de prédiction en cas de couverture de marqueurs insuffisante. Pour la sélection d'animaux de différentes races, le problème est encore plus important car les animaux partageront de plus petits segments et la prédiction de la valeur génétique génomique sera moins précise (Hayes *et al.*, 2013).

Il est donc important d'identifier le plus de SNP informatifs possibles afin de sélectionner ceux qui permettront de capter la plus grande variance génétique possible (Hayes *et al.*, 2013). Différentes approches ont été développées (Goddard et Hayes, 2009). Tout d'abord, le séquençage génome entier est une option



mais elle est encore très coûteuse. Il peut néanmoins être fait sur les animaux fondateurs de chaque lignée pour sélectionner des SNP et créer une puce faible densité sur laquelle imputer ensuite les génotypes en génome entier. L'approche gène candidat est limitée par la connaissance des gènes impliqués dans les mécanismes biologiques de régulation des phénotypes d'intérêt. Il faut ensuite identifier les SNP qui peuvent altérer ce gène.

1 Le génome du porc et les puces de génotypage disponibles

Le premier assemblage de référence du génome du porc domestique *Sus scrofa* v10.2 a été publié en 2012 par le *Swine Genome Sequencing Consortium* (Groenen *et al.*, 2012), dérivé d'une femelle Duroc. Le génome porcin comporte 18 paires de chromosomes et une paire de chromosomes sexuels XY. Une nouvelle version du génome porcin a été publiée en août 2017 (*Sus scrofa* v11.1). D'après la base de données Ensembl (release 90), 22 452 gènes ont été annotés dans le génome porcin.

Actuellement, 67 millions de SNP sont référencés et 195 millions ont été soumis à la base de données internationale dbSNP (Build 150) (<https://www.ncbi.nlm.nih.gov/projects/SNP/>). Des puces de génotypage à moyenne et haute densité sont disponibles. La puce Illumina PorcineSNP60 (Ramos *et al.*, 2009) comporte aujourd'hui 64 232 SNP identifiés grâce au séquençage génome entier de 18 individus (Duroc, Landrace, Large White, Pietrain, Sanglier).

Plus récemment, Affymetrix a produit une puce très haute densité avec 658 692 SNP : Axiom Porcine Genotyping Array (Axiom_PigHDv1) (<https://www.thermofisher.com/order/catalog/product/550588>). La construction de cette puce a reposé sur le séquençage génome entier de 210 animaux, issus de races commerciales ou non, choisies pour leur intérêt pour les programmes d'élevage.

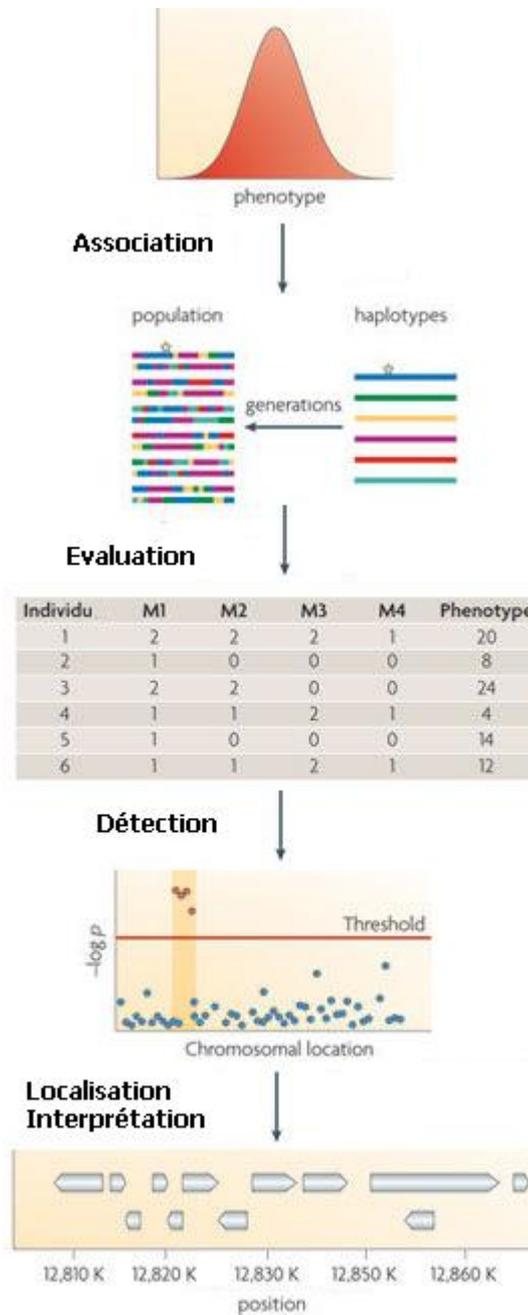
2 Etude d'associations génome entier (GWAS)

a) Principe

La génétique d'association à l'échelle du génome (GWAS pour *Genome Wide Association Study* en anglais) teste des associations statistiques entre des polymorphismes et la variabilité d'un caractère quantitatif sur un ensemble d'individus *a priori* non apparentés (Goddard et Hayes, 2009; Mackay *et al.*, 2009). Pour chaque SNP, on étudie la variation d'un caractère quantitatif en fonction du génotype du SNP pouvant prendre trois valeurs (AA, Aa, aa). Les différences de moyennes entre classes homozygotes au marqueur donnent une estimation de l'effet additif. Le contraste entre la classe hétérozygote et la moyenne des classes homozygotes permet une estimation de l'effet de dominance. Le plus souvent en génétique animale,



l'association est estimée par un modèle linéaire, un modèle de régression ou des modèles bayésiens, incluant l'effet du SNP comme effet fixe et la valeur polygénique (Goddard et Hayes, 2009). Le Manhattan plot permet de visualiser les régions présentant une association entre génotype et phénotype (Figure I-14).



Nature Reviews | Genetics

Figure I-14 : Analyse d'association génome entier (Figure traduite et adaptée et Mackay *et al.*, 2009)

L'utilisation d'une forte densité de marqueurs permet également de réduire l'intervalle de confiance de la localisation du QTL. Les QTL identifiés sont supposés être en déséquilibre de liaison avec les mutations causales des variations des caractères associés. La puissance de détection d'un QTL du test d'association



dépend de différents facteurs (Hayes, 2013) : (i) le déséquilibre de liaison entre le marqueur et le QTL ; (ii) la proportion de la variance du phénotype expliquée par le QTL ; (iii) la taille de la population étudiée ; (iv) la fréquence allélique de l'allèle le plus rare ; (v) le niveau de significativité alpha fixé par l'utilisateur ; (vi) la taille de la population ; (vii) la stratification de la population. Pour prendre en compte la structure familiale souvent présente dans les populations d'élevage, il faut corriger par la structure familiale (pedigree ou matrice de relation génomique) car elle explique souvent une partie de la variabilité phénotypique (Würschum, 2012). Le développement des marqueurs moléculaires génétiques, notamment des SNP, a permis d'estimer les relations d'apparentement entre les animaux sous la forme d'une matrice génomique, en s'affranchissant des pedigrees qui peuvent être erronés, sous condition d'une forte densité de SNP (Lopes *et al.*, 2013). Il a d'ailleurs été montré que la variation de relation entre individus apparentés est mieux estimée quand on utilise l'information génomique plutôt que le pedigree.

Les études génomes entiers impliquent de travailler sur des milliers de marqueurs génétiques avec une multitude de tests statistiques indépendants. En général, un seuil de significativité arbitraire est fixé pour chaque test. Cependant, le seuil global de l'expérience ne correspond plus au seuil de significativité choisi au départ du fait de l'accumulation du taux de faux positifs inhérents à chaque test. Il faut donc appliquer une correction dite de tests multiples (Hayes, 2013). Il en existe plusieurs qui sont plus ou moins strictes et plus ou moins adaptées aux approches génétiques. Les plus utilisées sont (Hayes, 2013; Laird et Lange, 2006) :

- La correction de Bonferroni : elle ne prend pas en compte le fait que les tests pour un même chromosome peuvent ne pas être indépendants (en déséquilibre de liaison). Elle est très conservative et pénalise les régions avec des SNP très corrélés ;
- Les tests de permutations : ils réattribuent de manière aléatoire les phénotypes parmi les individus pour supprimer toute relation existante entre génotype et phénotype. Cette correction tient compte du déséquilibre de liaison. En général, cela ne permet pas de conserver l'effet pedigree s'il en existe un et cette correction limite le maximum de significativité par le nombre de permutations effectuées ;
- Le taux de faux positif ou False Discovery Rate (FDR): il détermine la proportion attendue de QTL détectés comme faux positifs en fonction du seuil de significativité, du nombre de QTL dépassant ce seuil et du nombre de marqueurs testés. Quand le seuil est moins stringent, le nombre de SNP significatifs augmente et le FDR également.

Les approches GWAS identifient le plus souvent des QTL avec un effet faible sur le trait, surtout quand il est complexe (Würschum, 2012). Des études complémentaires sont nécessaires pour identifier au sein de



ces loci les gènes et les QTN (pour *Quantitative Trait Nucleotide*), équivalents des mutations causales pour les traits (maladies) mendéliens.

b) QTL identifiés chez le porc

La recherche des QTL chez le porc a pris de l'ampleur ces 20 dernières années. La première publication de QTL chez le porc date de 1994 (Andersson *et al.*, 1994). De nombreux QTL ont ensuite été mis en évidence chez le porc mais peu de gènes ou de variants causaux ont à ce jour été identifiés. D'après la base de données PigQTLdb (Version 33 du mois d'août 2017 ; animalgenome.org), 25 610 QTL ont été publiés dans 593 articles. Ils représentent 646 traits différents. La Figure I-15 résume le nombre de QTL publiés par types de traits phénotypiques.

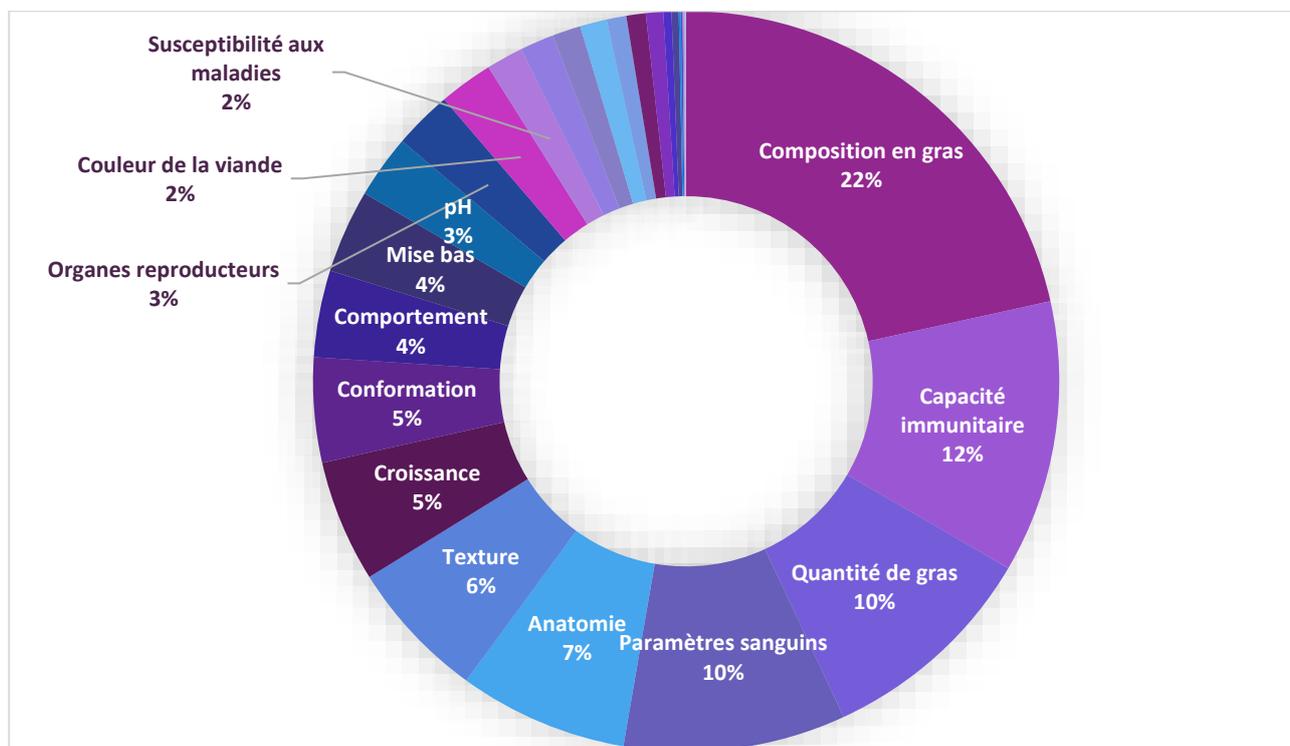


Figure I-15 : Nombre de QTL associés à des traits phénotypiques

c) Validation et identification des QTN

Les résultats obtenus dans les GWAS permettent idéalement d'établir une liste limitée de marqueurs pour des expériences de validation fonctionnelle (complémentation par transgénèse, analyse de mutants...). Avant la validation fonctionnelle souvent lourde à mettre en place, il est possible de réaliser une première



étape de validation, dite de validation génétique. Il s'agit de chercher à identifier la même association par d'autres approches et/ou dans d'autres groupe d'animaux et d'autres populations. Ainsi, il est possible de tester l'association en géotypant précisément le marqueur identifié par différentes méthodes dont le Pyroséquençage, la PCR allèle spécifique, le RFLP ou séquençage), dans un groupe d'animaux représentant les trois géotypes et de comparer les valeurs moyennes du trait associé de chaque groupe. Pour ce type de validation, les populations d'élevage présentent l'avantage de disposer de populations indépendantes issues d'autres cycles de production. Cela permet donc une validation des QTL et une élimination des faux positifs (Würschum, 2012).

Une validation génétique peut également être menée sur d'autres races, lignées ou familles. Si elles partagent le même phénotype et la même association, le QTL est validé. Si elles ne partagent pas le même phénotype mais que l'association est retrouvée, le QTL est invalidé. Cela peut signifier qu'il n'est pas causal ou qu'il s'agit d'un effet d'épistasie (Mackay, 2014) où l'effet d'un QTL sur le trait dépend d'un autre QTL, qui serait dans ce cas, absent de la seconde population.

Parfois, le QTL est absent dans d'autres lignées ou d'autres familles (Goddard et Hayes, 2009), résultat de la sélection naturelle, de l'évolution ou plus probablement de la pression de sélection de la production qui conduit à fixer des allèles en populations. Il serait ainsi intéressant de travailler sur des populations mixtes, regroupant plusieurs espèces pour faciliter la validation dans diverses populations par la suite, en incluant ce nouveau paramètre dans le modèle.

Une fois un QTL cartographié et validé, il est possible d'y chercher un gène candidat qui serait impliqué dans la variation phénotypique par une ou plusieurs mutations causales. Il faut établir la liste des gènes présents dans le QTL et physiologiquement cohérents avec le trait et les séquencer. Par exemple, chez le porc, Ernst et Steibel (2013) ont pu lister 23 gènes candidats publiés dans des associations avec des traits de production.

2. La génétique génomique ou cartographie du contrôle génétique du transcriptome

Comme présenté précédemment, le transcriptome du sang est largement utilisé comme phénotype intermédiaire de caractères immunitaires. L'émergence de la génétique génomique qui relie la variation de l'expression génique à des variations génétiques avec des approches dites eGWAS (*expression GWAS*), permet d'envisager l'étude du contrôle génétique du transcriptome sanguin en appui à l'étude de l'architecture des caractères immunitaires.



En effet, il a été établi que les changements de régulation d'expression des gènes contribuent souvent aux adaptations spécifiques des espèces, ainsi qu'aux variations inter-espèces de phénotypes complexes et à la variation inter-individuelle dans la susceptibilité aux maladies (Pai *et al.*, 2015). Ainsi, l'existence de variation naturelle dans la régulation des gènes pourrait être liée à des phénotypes complexes au sein et entre les espèces. De plus, le phénotype de variation d'expression est le phénotype le plus immédiat après les variations de séquences dans l'ADN et la route entre le génome et le phénotype passe par de phénotype d'expression (Rockman et Kruglyak, 2006). La variation quantitative des niveaux d'expression de gènes constitue donc comme un phénotype intermédiaire entre l'ADN génomique et la variation de séquence et des phénotypes plus complexes au niveau cellulaire, organe ou organisme (Petretto *et al.*, 2006). À cause des changements physiologiques dans les types cellulaires spécifiques avec une maladie ou un phénotype, il apparaît logique que les corrélations entre le locus associé à ces changements physiologiques et l'expression des gènes puissent être détectables dans les tissus appropriés (Min *et al.*, 2011).

Nous allons donc présenter dans cette partie des méthodes de mesures de l'expression des gènes ainsi que des méthodes d'analyses de génétique génomique.

a) La complexité du transcriptome

Le dogme de la biologie établi par Crick (1958, 1970) décrivait que l'information génétique était stockée dans les gènes sous forme d'ADN, puis transcrit en ARN et enfin traduit en protéines (Figure I-16). Cependant, il est désormais clair que, chez l'Homme comme chez le porc, les régions codant pour des protéines (appelées gènes) ne représentent que 2% de l'ADN génomique et qu'une partie significative du génome est transcrit sous forme d'ARN non codants (Mattick, 2001).



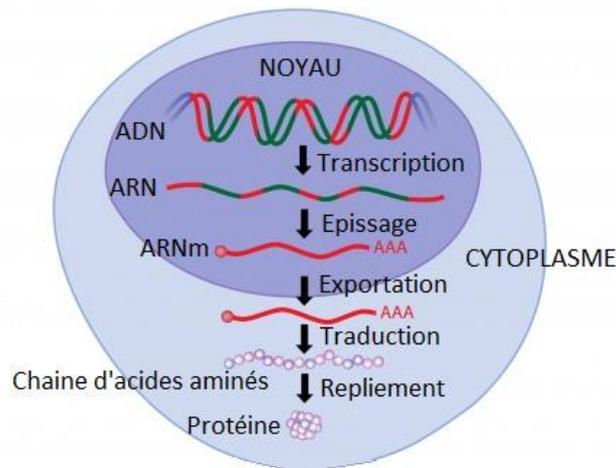


Figure I-16 : Représentation canonique du processus ADN-ARN-protéine (source : Figure traduite de Alila Medical Media's)

Au départ considérés comme de simples intermédiaires encapsulés dans le dogme ADN-ARN-protéine, les ARN non codants sont désormais envisagés comme des acteurs de la complexité et de la régulation de l'expression des génomes chez les eucaryotes (Mattick, 2001). Ainsi, même si l'architecture du contrôle de l'expression des gènes est en premier lieu en *cis* du gène par l'action des promoteurs et des activateurs et secondairement en *trans* par l'action de facteurs de transcription, les ARN non codants rajoutent un niveau de complexité (Mattick, 2001). Certains sont issus de la transcription d'introns et d'autres des régions intergéniques. Les ARN ribosomiaux (ARNr) et les ARN de transfert (ARNt) sont impliqués dans la traduction des ARN messagers (ARNm). Les ARN nucléaires (ARNsn) ou nucléolaires (ARNsno) sont impliqués respectivement dans l'épissage et la modification des ARNr (Mattick et Makunin, 2006). De petits ARN non codants sont également impliqués dans la régulation de l'expression des gènes après la transcription (étape post-transcriptionnelle) comme les microARN et les piARN (*piwi-interacting* ARN) (Stefani et Slack, 2008). Les ARN long non codants ont de nombreuses fonctions dont le remodelage de la chromatine, le contrôle de la transcription, des modifications post-transcriptionnelles (Mercer *et al.*, 2009) (Figure I-17).

Certains ARN longs codants sont issus de la transcription antisens d'un gène (Kukurba and Montgomery, 2015). Ceux-ci peuvent constituer des éléments de régulation négatifs de la transcription des gènes par interférence, des mécanismes post-transcriptionnels ou silencing par RISC (He *et al.*, 2008; Katayama *et al.*, 2005). Ces ARN ont parfois été associés à des maladies chez l'homme, ou comme ayant un rôle dans le développement et l'adaptation à l'environnement, notamment par l'altération des profils d'expression, plus fréquente chez les eucaryotes que chez les procaryotes (Mills *et al.*, 2013). Chez le porc,



Chen *et al.* (2012) ont étudié la transcription par séquençage d'ARN sens et antisens dans le foie au sein d'une génération F2. Ils ont observé que près de 30% des transcrits pour le muscle et 50% pour le foie présentent une transcription antisens.

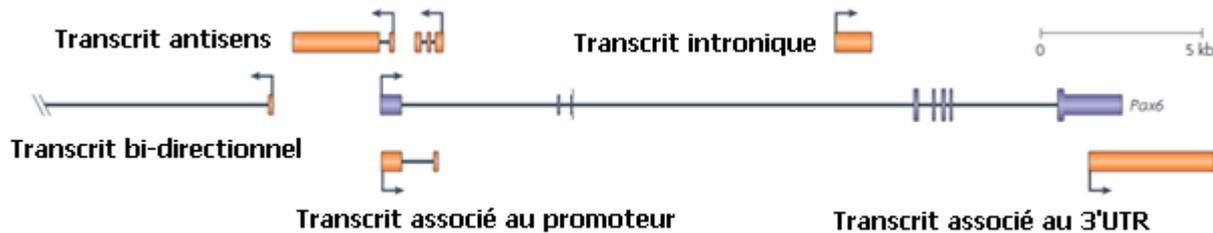


Figure I-17 : Organisation génomique des transcrits codants et non codants (Figure traduite de Kukurba et Mongemory, 2015)

L'expression des gènes se mesure en terme de quantité de transcrits. A petite échelle, la quantification de l'expression des gènes peut se faire par qPCR ou PCR quantitative après une étape de reverse transcription pour obtenir des molécules d'ADN complémentaires (ADNc). Cette méthode est précise, rapide et sensible mais elle ne peut être appliquée que sur un faible nombre de gènes et d'échantillons (Schroyen et Tuggle, 2015). Des technologies telles que la puce Fluidigm digital PCR a permis le développement de la PCR quantitative haut débit (96x96), nécessitant moins d'ARN et donc avec une plus grande sensibilité aux gènes faiblement exprimés (Schroyen et Tuggle, 2015). A très haut débit, les deux technologies principalement utilisées sont la technique de puce d'hybridation et le séquençage d'ARN (RNA-Seq).

- Hybridation de puces à expression

Utilisée pour la première fois en 1995 (Schena *et al.*, 1995), le principe de cette technologie repose sur l'utilisation de sondes (petites séquences d'oligonucléotides simple brin) complémentaires aux ARN transcrits, fixées au départ sur une membrane de nylon. Puis des puces sur de plus petites lames de verre ont été développées, permettant le dépôt d'une plus importante densité de sondes. Plus récemment, ce sont des puces dites à oligonucléotides sur lesquelles les sondes sont synthétisées *in situ* qui sont utilisées. Les puces d'expression les plus denses contiennent jusqu'à plusieurs centaines de milliers de sondes. Le protocole d'hybridation est présenté dans la Figure I-18.



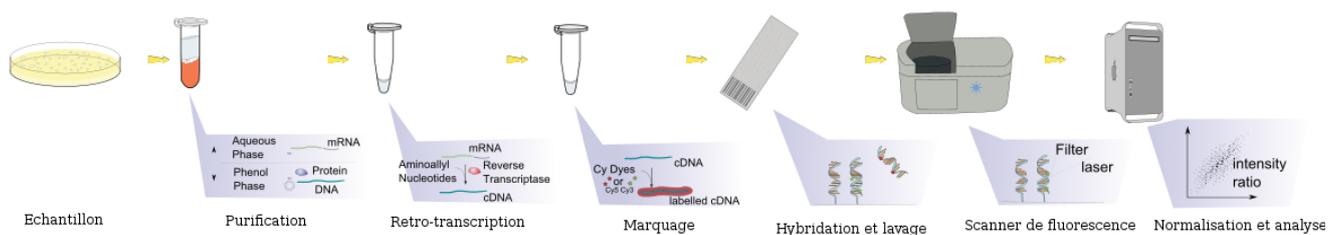


Figure I-18 : Protocole d'utilisation des puces à expression

Le design des sondes est réalisé à partir des séquences disponibles et permet de cibler des ensembles de gènes codants pour des protéines notamment. Les ARN extraits des échantillons sont marqués par un ou deux fluorochromes (Cyanine 3 ou Cyanine 5) et sont déposés sur les puces. Ils vont alors s'hybrider sur leurs sondes complémentaires, organisées en spot. Chez le porc, Gao *et al.* (2010) ont élaboré une puce SLA-RI/NRSP8-13K enrichie pour les gènes de fonctions liées à l'immunité et couvrant le CMH. Elle spécifie ainsi 10 010 gènes sans redondance. Une puce 60K Qiagen-NRSP8 a plus tard été développée et enrichie depuis notamment pour des gènes de l'immunité.

Les puces d'expression sont soumises à certains biais et des limites propres à cette technologie. L'étude de l'expression de gènes n'est limitée qu'aux sondes présentes sur la puce (Kukurba et Montgomery, 2015). De plus, le signal d'intensité de fluorescence mesuré est une combinaison du signal biologique et du bruit expérimental. Une étape de normalisation permet d'éliminer les variations non biologiques mais il reste parfois difficile de faire la différence entre la biologie et le bruit. Il y a également des artéfacts dûs à des hybridations croisées entre des séquences trop similaires (Casneuf *et al.*, 2007). Enfin, les puces à expression ne peuvent pas quantifier précisément le niveau de transcription de gènes faiblement ou très fortement exprimés (Kukurba et Montgomery, 2015).

L'intensité du signal pour un spot est proportionnelle à la quantité de transcrits correspondant à la sonde. Des étapes de traitements bioinformatiques sont nécessaires pour obtenir une évaluation de l'expression des gènes la plus précise possible :

- Il faut normaliser par l'intensité du bruit de fond, c'est à dire l'intensité de fluorescence mesurée même là où il n'y a pas de sonde. Ce bruit de fond peut varier d'une sonde à une autre donc il faut éliminer ce biais pour pouvoir comparer les échantillons.
- Il faut éliminer les signaux trop faibles, c'est-à-dire proche des témoins négatifs.
- Les données sont normalisées pour pouvoir comparer les puces et elles peuvent être transformées pour permettre plus tard l'utilisation d'approches statistiques paramétriques.



- Technologies de séquençage

Le développement des méthodes de séquençage haut-débit nouvelle génération (NGS pour *next generation sequencing*) a révolutionné l'étude du transcriptome (Costa *et al.*, 2010 ; Wang *et al.*, 2009). Le RNA-Seq permet une étude de l'expression des gènes, des épissages alternatifs et des expression allèles spécifiques (ASE pour *allele specific expression*). C'est le choix du protocole d'extraction des ARN qui permettra de cibler certaines classes d'ARN. En effet, comme montré sur la Figure I-19, le protocole standard d'un RNA-Seq comporte une étape de sélection des ARN, sur des critères de taille ou de présence de queue polyA spécifique des ARNm et ARN antisens.

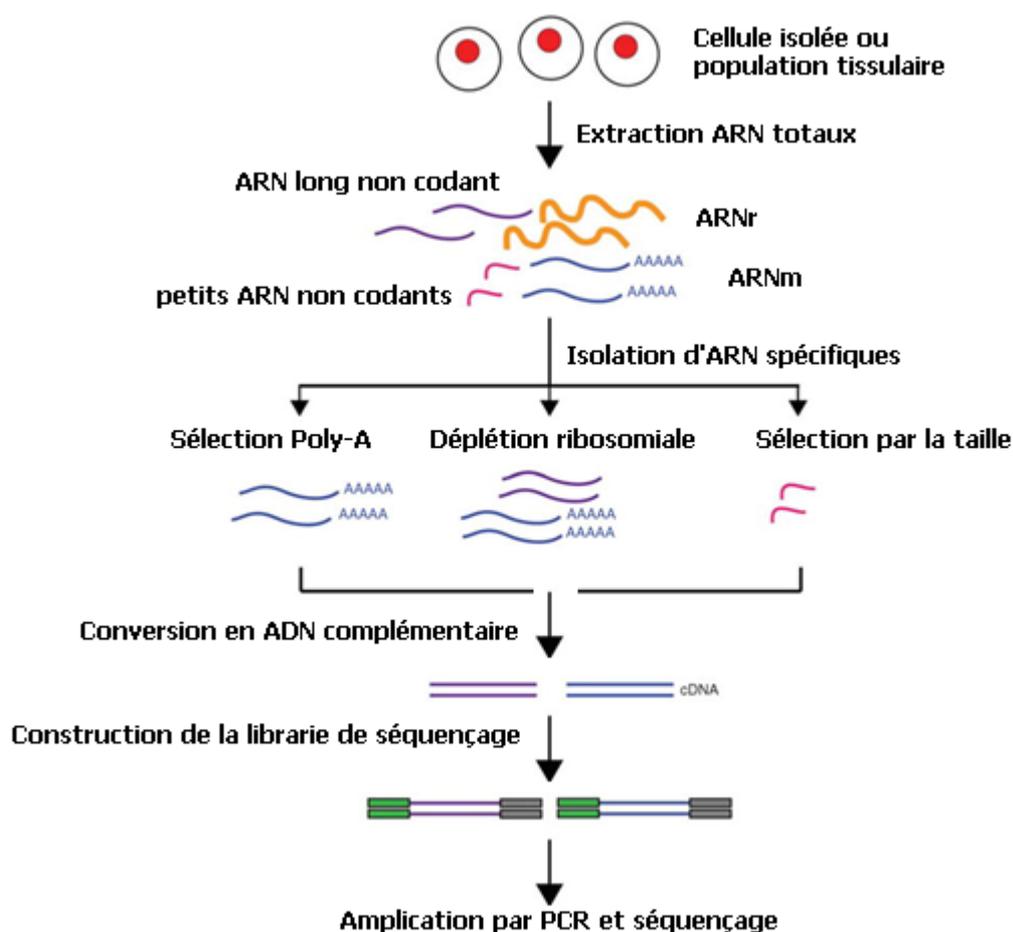


Figure I-19 : Protocole de séquençage d'ARN (Figure traduite de Mills *et al.*, 2013)

Des protocoles spécifiques de RNA-Seq dits « stranded » permettent désormais de distinguer les ARNm des ARN antisens. Ils retiennent l'information du brin d'ADN à partir duquel l'ARN a été transcrit (Mills



et al., 2013) : (i) Utilisation d'adaptateurs à orientation brin spécifique ; (ii) Marquage des ARN au bisulfite, ce qui entraîne un changement des cytidines (C) en uridines (U) et limite la fixation des séquences à un seul brin ; (iii) Etape de synthèse du 2^{ème} brin modifié par incorporation de dUTP lors de la reverse transcription. Ainsi, le brin original aura des résidus deoxythymidine.

Le séquençage des ARN présente de nombreux avantages par rapport aux technologies d'hybridation (Schroyen et Tuggle, 2015) : (i) il ne nécessite ni primers ni sondes spécifiques de gènes ; (ii) le bruit de fond est faible, voire inexistant ; (iii) il n'y a pas nécessité de connaissance préalable des séquences, c'est donc une méthode intéressante pour les espèces d'élevage avec un génome mal annoté, ou sans séquence de référence ; (iv) il permet la détection des sites d'initiation de la transcription (TSS), de variants d'épissage, d'ARN non codants et de transcrits antisens ; (v) il permet l'identification de nouveaux gènes ou de transcrits à expression faible si la profondeur de couverture est suffisante ; (vi) l'échelle de détection est plus étendue.

Toutefois, le séquençage d'ARN est coûteux (de cinq à six fois le prix d'une puce), même si son prix a bien diminué ces dernières années. De plus, il faut parfois envisager la déplétion de gènes très abondants de faible intérêt comme l'alpha et la beta globine dans le sang avant le séquençage pour améliorer la sensibilité de détection.

La Figure I-20 présente le protocole standard d'analyse de données RNA-Seq. La sortie des séquenceurs consiste en la séquence en bases de chaque fragment ou lecture dans un fichier de type « .fastq ». Pour tout type d'analyse, la première étape consiste en un contrôle qualité des lectures pour vérifier notamment la qualité des bases, la longueur des lectures, les erreurs de séquençage. Des étapes de traitements intermédiaires améliorent la qualité des lectures. Les lectures avec une qualité globale faible sont éliminées. Le *trimming* permet de couper les lectures aux extrémités si la qualité des bases aux extrémités est faible. On peut également enlever les séquences des adaptateurs utilisés lors du séquençage.

Puis les lectures sont alignées sur le génome de référence ou peuvent être soumises à un assemblage *de novo* (c'est-à-dire sans assemblage de référence) (Kukurba et Montgomery, 2015). La particularité des données de séquençage d'ARN est liée à l'absence des introns. Seuls les exons ont été séquencés alors que le génome de référence est composé des introns et des exons. Il est donc possible que les lectures de séquençage d'ARN représentent une séquence qui s'étend de part et d'autre d'une jonction d'épissage. Des outils d'alignements tels que TopHat2 (Trapnell *et al.*, 2009) et STAR (Dobin *et al.*, 2013) permettent de tenir compte de ces jonctions.



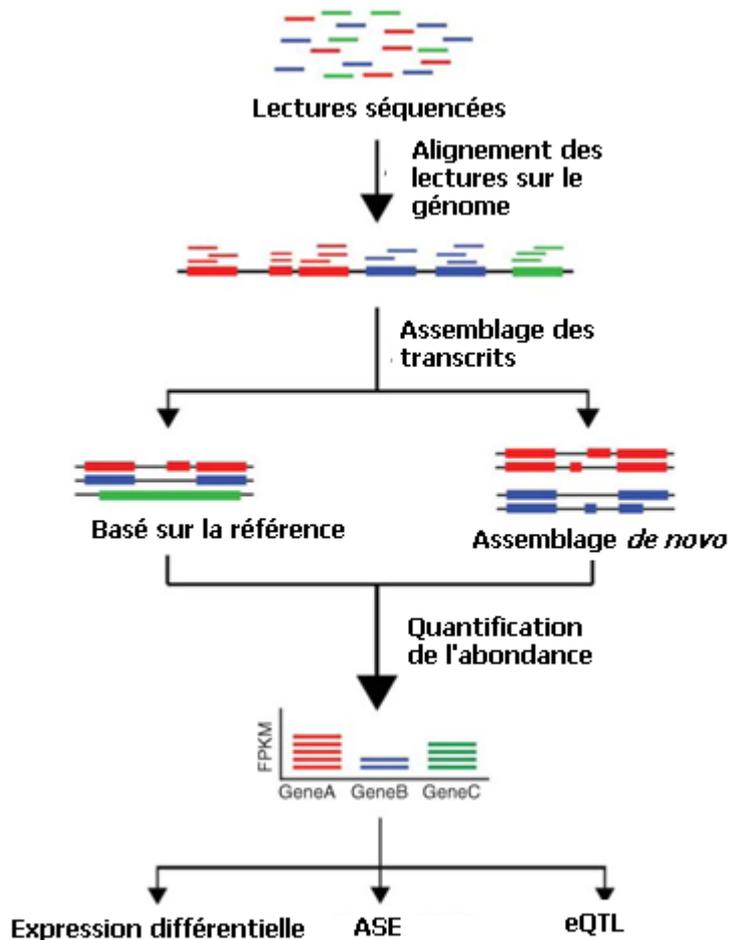


Figure I-20 : Protocole d'analyse standard de données RNA-Seq (Figure traduite de Mills *et al.*, 2003)

b) Etude du contrôle génétique de l'expression

Il a été montré que l'expression des gènes est héritable. Elle a été estimée pour la première fois chez *S. cerevisiae* dans un croisement entre des souches expérimentales et sauvages (Brem *et al.*, 2002). Par la suite, elle a pu être estimée notamment chez le rat (Petretto *et al.*, 2006) ou chez l'Homme (Emilsson *et al.*, 2008 ; Göring *et al.*, 2007). Il semble donc qu'une part de la variabilité de l'expression des gènes est imputable à la variation génétique et qu'il est envisageable de cartographier le contrôle génétique du transcriptome. Dès 2001, Jansen et Nep avaient ainsi proposé pour la première fois une analyse combinant cartographie génétique et technologie de puce d'hybridation transcriptomique posant ainsi les bases de la génétique génomique.



- eGWAS

Les études d'association génome entier pour l'identification de loci associés aux niveaux d'expression de gènes, considérés alors comme un trait quantitatif, sont connues sous le terme eGWAS. Cette approche intègre les profils de transcription à des données de génotypage en considérant les niveaux d'expression comme des réponses variables afin d'identifier des QTL associés à l'expression ou eQTL (Ernst et Steibel, 2013). Les eGWAS peuvent être appliquées à des données de puces transcriptomiques ou des données de séquençage d'ARN.

Des eGWAS ont été réalisées dans différentes espèces, sur différents tissus avec entre 20% et 30% de la variance de l'expression des gènes expliquée par des variations génétiques : Brem *et al.* (2005) chez la levure, Schadt *et al.* (2003) chez la souris pour la première eGWAS chez les mammifères, Morley *et al.* (2004) chez l'Homme. Les études GWAS publiées chez le porc sont résumées dans le Tableau I-6.

Tableau I-6 : Sélection d'études eGWAS publiées chez le porc

Ponsuksili <i>et al.</i> (2010)	Muscle long dorsal	9 180 eQTL
Liaubet <i>et al.</i> (2011)	Muscle squelettique	335 eQTL
Steibel <i>et al.</i> (2011)	Muscle (loin)	62 eQTL
Cinar <i>et al.</i> (2012)	Muscle long dorsal	Pour 3 gènes identification d'un SNP
Chen <i>et al.</i> (2013)	Foie	120 <i>cis</i> -eQTL et 523 <i>trans</i> -eQTL
Reiner <i>et al.</i> (2014)	Poumon	55 eQTL
Ponsuksili <i>et al.</i> (2014)	Muscle long dorsal	47 836 eQTL
Kogelman <i>et al.</i> , (2015)	Tissu adipeux	987 <i>cis</i> -eQTL et 73 <i>trans</i> -eQTL
Martinez-Montes <i>et al.</i> (2016)	Muscle long dorsal	63 eQTL
Gonzalez-Prendes <i>et al.</i> (2016)	Muscle long dorsal	4 eQTL

Une eGWAS repose sur le même principe statistique qu'une GWAS classique considérant l'expression de chaque gène comme un phénotype quantitatif indépendant et comme pour les GWAS il faut tenir compte de la structure familiale et corriger pour les tests multiples (Kendzioriski et Wang, 2006). Les études eQTLs contribuent à la compréhension des cinq questions cruciales en génétique génomique en partie décrites par William *et al.* (2007) :

- Quelle proportion de gènes avec une expression sous contrôle génétique ?

En fonction des études, la proportion de traits d'expression liés ou associés à au moins un eQTL s'étend de 0,8 à 59 %. En moyenne, un eQTL est détecté pour 10 à 30 % des gènes testés. La puissance statistique des études est cruciale et dépend principalement de la taille de l'échantillon. Multiplier la taille de l'échantillon par deux peut augmenter par quatre la puissance de détection des eQTL (Williams *et al.*, 2007).



- De quelle nature est l'influence génétique sur l'expression des gènes ?

La variation génétique influençant l'expression des gènes peut être dans une séquence régulatrice comme un promoteur, un activateur, des sites d'épissage, des motifs secondaires de gènes cibles. Elle est alors caractérisée comme agissant en *cis* et définit ainsi un eQTL *cis* (Haldane, 1941). Les eQTL *cis* agissent de façon allèle spécifique sur l'expression du gène. C'est le cas pour des variants situés dans des régions promotrices qui vont moduler l'initiation de la transcription, les variants qui modifient la structure de la chromatine, les variants en 3' qui concentrent les sites de fixation de microARN influençant la stabilité de l'ARNm, les variants modifiant les sites d'épissage ou les variants dans la région exonique qui vont entraîner la production d'une protéine autorégulatrice (Figure I-21) (Skelly *et al.*, 2009). La variation génétique peut également entraîner une variation dans la séquence de protéines (facteurs de transcription) ou d'ARN (microARN) interagissant avec une séquence régulatrice de la transcription du gènes (Figure I-21). Elle est alors qualifiée comme agissant en *trans* (eQTL *trans*) et affectent l'expression des deux allèles d'un gène.

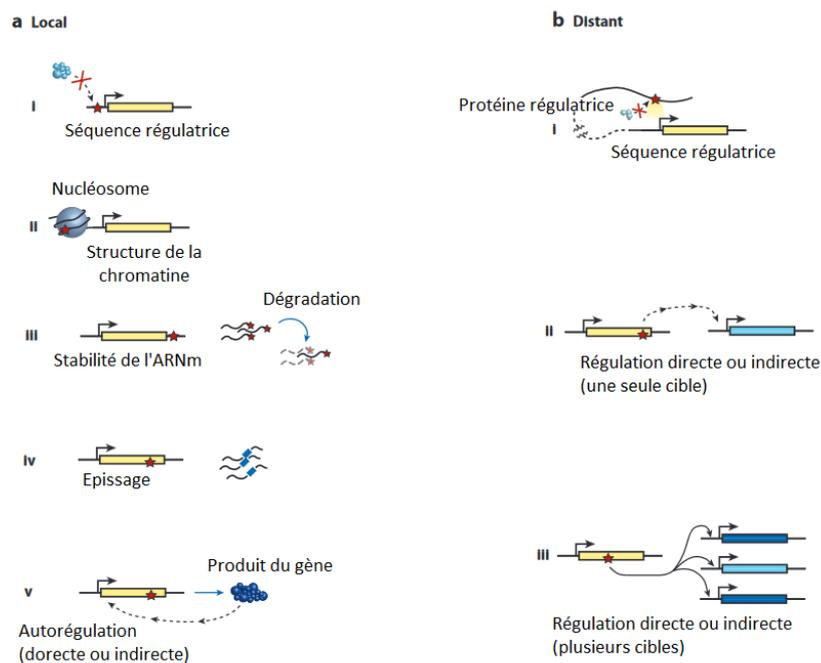


Figure I-21 : Mécanismes moléculaire des variants régulateurs *cis*/local et *trans*/distant (Figure traduite de Skelly *et al.* 2009)

Les eGWAS permettent la détection des effets *cis* et *trans* mais une ambiguïté persiste sur la caractérisation précise des effets. Dans de nombreuses études, il a été défini une distance génomique arbitraire caractérisant les eQTL les plus proches du gène affecté comme eQTL *cis* et les plus éloignés, comme eQTL *trans*. Les variants *trans* sont d'ailleurs plus souvent sur d'autres chromosomes (Pai *et al.*, 2015). Or il a été montré que des variants agissant en *cis* peuvent être à distance (jusqu'à des centaines de kilobases) du



gène affecté (Smith *et al.*, 2013) et que des variants *trans* peuvent en être proches. Pour pallier ce problème, Rockman et Kruglyak (2006) ont fait une distinction local/distant au lieu de *cis/trans*.

Les effets *cis* sont plus forts et les plus grandes différences d'abondance de transcrits ont été associées à des variations *cis*. Ils sont donc plus facilement détectables car ils expliquent une plus large fraction de la variance de l'expression. La proportion de eQTL *cis* augmente avec la taille de la fenêtre et le seuil de significativité. La valeur de l'héritabilité estimée est supérieure pour les eQTL *cis* par rapport aux eQTL *trans*. L'utilisation de seuil de significativité stringent entraîne de nombreux faux négatifs, notamment parmi les signaux *trans*. Les études d'héritabilité suggèrent toutefois que plus de la moitié de la variance de l'expression est due à des eQTL (Pai *et al.*, 2015).

- Existent-ils des polymorphismes génétiques associés à l'expression de nombreux gènes ?

Dans le cas où une protéine (comme un facteur de transcription) contrôle l'expression de nombreux gènes, la variation génétique affectant son expression (en *cis*) peut avoir une influence sur ses gènes cibles (en *trans*). Cela forme un régulon et la variation génétique *trans* peut être qualifiée de régulatrice majeure (*master regulator* en anglais). La corrélation de variation de ces groupes de gènes peut avoir des conséquences sur le phénotype. Dans les différentes études, jusqu'à 17 régulateurs majeurs ont été trouvés (Williams *et al.*, 2007).

- L'expression des gènes est-elle soumise à des contrôles génétiques multiples ?

Une approche de détection d'eQTL analyse des milliers de traits d'expression. Cela entraîne une importante complexité statistique et des besoins considérables en capacité de calculs. Certaines études ont choisi une approche simple locus pour limiter la complexité. Mais cela empêche l'étude de régulateurs multiples de l'expression (Williams *et al.*, 2007). Pourtant des études génome entier ont identifié des traits d'expression associés à différents eQTL (Brem *et al.*, 2002).

- Quelle est la spécificité tissulaire du contrôle génétique des gènes ?

L'augmentation de la variabilité biologique dans des tissus complexes et hétérogènes ou la susceptibilité de la variation de transcription aux facteurs environnementaux peut expliquer une distribution différente du nombre de signaux *trans* détectés par Petretto *et al.* (2006) dans les tissus rénaux et de la glande surrénale. Les études sur le sang ont publié le même type d'observation. Comme les eQTL *trans* sont plus polygéniques, les profils de signaux *trans* détectés sont plus souvent liés à la régulation spécifique au tissu avec une limitation des *trans* détectables dans différentes expériences (Petretto *et al.*, 2006). Selon Ackermann *et al.* (2013), 5 % à 94 % des eQTLs sont spécifiques des cellules en fonction de l'étude. Mais la spécificité cellulaire est certainement surestimée à cause de la faible puissance de détection des études eQTL



dans certains types cellulaires. Elle est en effet soumise au dynamisme des processus cellulaires comme par exemple la différenciation cellulaire ou la réponse à des stimuli.

Lors des analyses eGWAS, il faut bien maîtriser la balance entre puissance statistique et généralisation. Plus les échantillons seront spécifiques et homogènes, plus la puissance de détection sera importante (lignées cellulaires, organismes modèles) mais cela diminue par ailleurs la généralisation des résultats à d'autres cellules/tissus/conditions/espèces. Cela est d'autant plus vrai pour les associations avec des variants distaux agissant en *trans*, plus difficile à détecter du fait d'un effet plus faible (Gilad *et al.*, 2008). Il est d'ailleurs difficile en général de comparer les différentes études eQTL car les seuils de détections, les méthodes statistiques, les protocoles d'échantillonnage et donc la puissance de détection sont souvent très différents. Cette faible répétabilité des résultats complique la validation des eQTL notamment *trans*. Cependant, l'identification d'un effet *cis* pour un SNP décrit avec un effet *trans* sur un autre gène pourrait valider cet eQTL (Pierce *et al.*, 2014). Ce type de variant permet de mettre en évidence de potentiels mécanismes de régulation.

Comme décrit précédemment, le sang est un tissu complexe contenant de multiples types cellulaires et autres éléments circulants. Ainsi, la cartographie des eQTL pour le transcriptome du sang total sera sous l'influence de la composition sanguine (Schramm *et al.*, 2014). Plusieurs études ont choisi de cartographier les eQTL dans des sous population cellulaires sanguines (Min *et al.*, 2011; Pierce *et al.*, 2014; Rotival *et al.*, 2011)

- ASE

Dans l'ADN génomique d'un tissu, le ratio de représentation des allèles présents sur chaque copie de chromosome est 50:50. Le développement normal et les processus cellulaires peuvent pourtant entraîner un ratio d'expression des allèles différent, ou une balance d'expression allélique (Figure I-22). Il peut s'agir de l'effet d'une influence génétique agissant en *cis* sur l'expression du gène ou une empreinte parentale, phénomène épigénétique entraînant une différence de niveau d'expression entre les deux allèles parentaux avec une tendance prononcée pour l'une ou l'autre origine (plus fréquemment pour le père) avec par exemple l'inactivation du X chez les femelles (Baran *et al.*, 2015). Une balance allélique peut également être le résultat d'un variant non-sens entraînant l'apparition prématurée d'un codon stop pour une des copies du gène (MacArthur *et al.*, 2012).



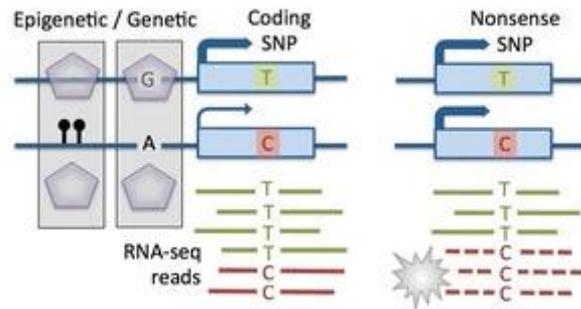


Figure I-22 :Représentation d'effet ASE détecté par séquençage (source : Castel *et al.*, 2015)

Grâce aux données de séquençage d'ARN, il est possible d'étudier cette différence d'expression allélique de variants codants dans des individus hétérozygotes (Castel *et al.*, 2015). Ces approches permettent de compléter les cartographies d'eQTL obtenues par eGWAS en caractérisant directement les variants agissant en *cis*. La puissance de ces analyses repose sur un meilleur contrôle intra-individuel éliminant ainsi les facteurs environnementaux et les influences génétiques agissant en *trans* (Pastinen, 2010). Les premières analyses ASE indiquent que les différences allèle-spécifiques entre les transcrits chez un individu peuvent affecter jusqu'à 30 % des loci, et au niveau de la population, environ 30 % des gènes exprimés montrent que leur régulation en *cis* est influencée par des allèles communs (Pastinen, 2010).

L'utilisation de données de séquençage implique une première étape d'alignement des lectures avec les aligneurs précédemment présentés (TopHat2 (Trapnell *et al.*, 2009), STAR (Dobin *et al.*, 2013)). Puis il faut compter le nombre de lectures alignées spécifiquement sur chaque allèle des variants considérés. Le Broad Institute a par exemple implémenté l'outil ASEReadCounter (Van der Auwera *et al.*, 2013) qui va permettre d'obtenir pour chaque variant sa fréquence allélique par chaque individu.

Cependant, il existe un biais d'alignement sur le génome de référence du fait de la présence de polymorphismes dans les lectures. Pour détecter des expressions allèle-spécifiques, il est donc essentiel de contrôler ce biais. D'une part, il faut s'assurer d'avoir une importante profondeur de couverture afin de limiter l'impact des erreurs de séquençage. D'autre part, il faut limiter le biais d'alignement (Degner *et al.*, 2009; Satya *et al.*, 2012; Wood *et al.*, 2015) et plusieurs méthodes ont été proposées : (i) remplacer les SNP par une troisième base ou un "N" dans le génome de référence ; (ii) aligner sur le génome parental ; (iii) produire un génome de référence personnalisé incluant le génotype et l'haplotype de chaque individu (Wood *et al.*, 2015).



Il faut également pouvoir différencier les expressions soumises ou non à empreinte parentale (activation ou inactivation d'un allèle parental). Idéalement, il faut séquencer l'ADN de l'individu et des parents et estimer si par exemple l'allèle sur-exprimé provient toujours du même parent.

Statistiquement, la différence de lectures alignées sur l'allèle alternatif peut être modélisée par une distribution binomiale ou bêta-binomiale. Puis on réalise un test binomial sur les individus hétérozygotes (Castel *et al.*, 2015 ; Ernst et Steibel, 2013). Il est important de contrôler la qualité de l'alignement car des erreurs d'alignements, pour des lectures issues de régions similaires par exemple, vont entraîner la détection de faux ASE (Wood *et al.*, 2015). Il faut également éliminer les SNP contenus dans les insertions et les délétions.

De nombreuses études d'ASE sont publiées chez l'Homme (Larson *et al.*, 2015 ; Moyerbrailean *et al.*, 2016 ; Serre *et al.*, 2008; Smith *et al.*, 2013) notamment dans le programme GTEx (GTEx Consortium, 2013, 2015 ; GTEx Consortium *et al.*, 2017), mais aussi chez la souris (Crowley *et al.*, 2015 ; Hasin-Brumshtein *et al.*, 2014 ; Lagarrigue *et al.*, 2013). Les ASE chez les animaux domestiques sont aussi étudiés. Chamberlain *et al.* (2015) ont récemment publié une cartographie des ASE établie pour une vache et 18 tissus différents. Chez le porc, Yang *et al.* (2016) ont observé que 30% des gènes dans le muscle squelettique étaient influencés par une ASE et Wu *et al.* (2015) ont publié une analyse comparative d'ASE d'animaux infectés/non infectés.

3. Lier « expression de gènes-phénotypes » ou identification de biomarqueurs

On définit les marqueurs biologiques ou biomarqueurs comme des caractéristiques biologiques de processus pathologiques qui peuvent aider à prendre la décision d'interventions médicales pour prévenir le développement de maladies (Zhao et Qin, 2013) pour la recherche biomédicale. Pour la recherche appliquée aux animaux d'élevage, les biomarqueurs peuvent être compris comme des caractéristiques biologiques reflétant des capacités individuelles, pouvant être incluses dans le processus de sélection. Les biomarqueurs sont donc par définition des caractéristiques quantifiables des processus biologiques. En ce sens, la variation de l'expression d'un gène dans le sang peut constituer un biomarqueur. De plus, des eGWAS sur le sang total et notamment chez l'Homme (Schramm *et al.*, 2014) ont révélé un contrôle important du niveau d'expression des gènes dans le sang ce qui indique que le tissu sanguin complet est un tissu robuste pour les analyses eQTL et le confirme comme source de biomarqueurs d'expression intéressante.



L'approche la plus courante pour identifier des biomarqueurs candidats est une expression différentielle comparant de groupes d'animaux divergents (différentes caractéristiques biologiques, différentes conditions environnementales, différentes lignées) (Zhao et Qin, 2013). Il s'agit de repérer dans les gènes différentiellement exprimés entre les deux groupes. Ainsi, le niveau d'expression de ces gènes sera un biomarqueur pour identifier l'appartenance d'un animal à l'une ou l'autre des conditions étudiées. De nombreuses études d'expression différentielle ont été faites avec la technologies des puces à expression. Il est désormais courant d'étudier les différences d'expression à partir de données de RNA-Seq. Les modèles statistiques ont ainsi été adaptés. De nombreux outils ont été proposés et sont régulièrement comparés (Schurch *et al.*, 2016; Seyednasrollah *et al.*, 2015). Il semble ainsi que le choix de l'outil dépend du type d'étude menée (nombre d'échantillons, qualité des données, profondeur de couverture, précision recherchée).

Pour identifier des biomarqueurs au sein d'un seul groupe d'animaux, il faut combiner des données d'expression avec la variation phénotypique du trait d'intérêt. Par une approche univariée simple, chaque gène va être considéré comme indépendant et on estimera la corrélation (Pearson ou Spearman) ou l'association (modèle linéaire) de sa variation d'expression avec celle du trait. Des méthodes multivariées sont capables de prendre en considération la structure de corrélation de l'expression des gènes pour identifier la signature d'expression qui prédit le mieux la variation d'un phénotype, ou de plusieurs traits corrélés.

Les méthodes multivariées sont souvent associées à des procédures de sélection de variables. On distingue deux catégories de méthodes multivariées. La première regroupe des méthodes de détection non supervisée où l'évaluation de l'existence de groupes d'échantillons est suggérée par la méthode statistique elle-même. Parmi elles, la méthode la plus utilisée est l'ACP ou analyse en composante principale qui représente les objets décrits par les variables, dans un nouveau système de référence caractérisé par des combinaisons de variables appelées composantes principales. La première composante principale représente le maximum de variance contenu dans les données initiales et chaque composante principale suivante doit expliquer le maximum de la variance résiduelle. Ainsi, les variations systématiques sont expliquées par les premières composantes principales alors que le bruit expérimental et les variations aléatoires sont contenues dans les dernières composantes principales.

La deuxième catégorie comprend des méthodes de classification supervisée basée sur des informations *a priori* d'appartenance des échantillons à différentes classes. Pour cela, le principe de l'ACP a été adapté (*ranking PCA*). La méthode de régression des moindres carrés partiels (PLS pour *Partial Least*



Squares en anglais) est couramment utilisée. C'est une méthode de régression multivariée établissant une relation entre une ou plusieurs variables dépendantes (Y) et un groupe de descripteurs (X). Une relation de hiérarchie est établie au sein des groupes de variables.

Pour l'identification de biomarqueurs, il faut tenir compte des co-facteurs et il faut également un large échantillon, choisir un phénotype précis et une population bien caractérisée (Siemeling et Zeller, 2014). Il est important de caractériser l'efficacité d'un biomarqueur ou son utilité et de le valider. La validation de biomarqueurs nécessite une étude indépendante à grande échelle (Strimbu et Tavel, 2010).

4. Génotype-transcriptome-phénotype

Les études eQTL intègrent les profils de transcription aux données de génotypage en considérant les niveaux d'expression comme des réponses variables dans l'analyse des QTL pour identifier les gènes contrôlant des traits phénotypiques importants (Ernst et Steibel, 2013). Ainsi, les données de contrôle génétique de l'expression des gènes aident à l'interprétation des résultats de GWAS surtout quand elles sont issues de cellules ou tissus spécifiques du trait phénotypique d'intérêt.

La plupart des régions identifiées par GWAS correspondent soit à des déserts de gènes soit à des régions avec beaucoup de gènes tous aussi plausibles. Les données eGWAS peuvent ainsi aider à prioriser l'identification des gènes candidats dans ces loci.

En l'absence d'une étude eGWAS, et dans des espèces avec des bases de données riches comme l'humain ou la souris, il est de pratique courante après une GWAS de regarder dans les bases de données eQTLs pour déterminer si les SNP d'intérêt sont des eQTL pour des types de cellules intéressantes pour un phénotype d'intérêt. Quand les SNP ne sont pas des eQTL dans un type cellulaire d'intérêt, le lien s'arrête. Cependant il est possible que le type cellulaire n'ait pas été étudié, que la paire SNP-gène n'ait pas été étudié ou que la taille de l'échantillon soit trop faible pour une étude cellule spécifique (Brown *et al.*, 2013).

Quand les données de GWAS et eGWAS sont disponibles, l'intégration de leurs résultats consiste à rechercher des chevauchements significatifs entre ces deux types d'études, menées sur des cellules ou tissus en lien avec le phénotype d'intérêt. Si un gène avec des niveaux d'ARN messagers variables est trouvé dans un pic associé avec un trait physiologique, cela constitue une piste intéressante pour expliquer le mécanisme moléculaire et potentiel du gène candidat causal (Brown *et al.*, 2013; Conde *et al.*, 2013; Joehanes *et al.*, 2017; Lappalainen *et al.*, 2013; Ponsuksili *et al.*, 2015; Simon *et al.*, 2016; Williams *et al.*, 2007). Cela permet



de déduire un lien direct entre la variation du génotype et la variation du phénotype comme étant la différence d'expression entre les individus. C'est une méthode pour annoter la fonction d'un locus régulateur et pour identifier des variants causaux (Pai *et al.*, 2015).

Parfois, le chevauchement entre les loci identifiés en GWAS/eGWAS n'est qu'une coïncidence à cause de l'abondance des eQTL et du déséquilibre de liaison. Peñagaricano *et al.* (2015) ont donc proposé une méthode d'inférence de causalité dans les études de génétique génomique basée sur différentes étapes : (i) explorer les associations marginales entre le génotype et le phénotype (pQTL) par modèle linéaire et permutations ; (ii) explorer les associations entre le génotype et l'expression des gènes (eQTL) par modèle linéaire mixte et FDR ; (iii) chercher les régions génomiques où des pQTL co-localisent avec des eQTL ; (iv) inférer les réseaux gène-phénotype en utilisant des techniques de d'apprentissage de structures causales.

Ces méthodes prennent ainsi en compte la structure locale du déséquilibre de liaison et intègrent les eQTL et les résultats de GWAS. Nica *et al.* (2010) ont également proposé une solution pour éliminer les co-localisations accidentelles.

Si chez l'Homme il est essentiel de déterminer les variants causaux pour identifier l'origine génétique de maladies, pour les espèces d'élevage et dans un processus de sélection assistée par marqueurs, l'identification de marqueurs génétiques non causaux mais associés à un phénotype a déjà beaucoup de valeur. Ainsi, les SNP identifiés dans un eQTL ou avec un ASE pour un gène corrélé à un phénotype pourraient être inclus dans des processus de sélection s'ils permettent de prédire une part de la variabilité d'un trait d'intérêt.



Les objectifs de la thèse

Ce projet de thèse s'inscrit dans un programme de recherche conduit dans l'UMR GABI dont le but est de contribuer à identifier des leviers d'actions pour améliorer le statut sanitaire des élevages de porcs, afin de maîtriser la biosécurité des produits de consommation, le bien-être animal et les antibiorésistances tout en maintenant la productivité. Un des leviers envisagé est la sélection de reproducteurs qualifiés pour une capacité immunitaire favorisant la résistance aux pathogènes, l'efficacité vaccinale et la robustesse, limitant ainsi l'usage des anti-infectieux.

L'objectif global de ce travail de thèse est d'approfondir les connaissances sur l'architecture génétique des variations de paramètres immunitaires chez le porc, en intégrant des données de génotypage de SNP, de transcriptome et d'immunophénotypage (Figure O-23) issues des projets ANR IMMOPIG et SUS-FLORA (voir Annexe A1).

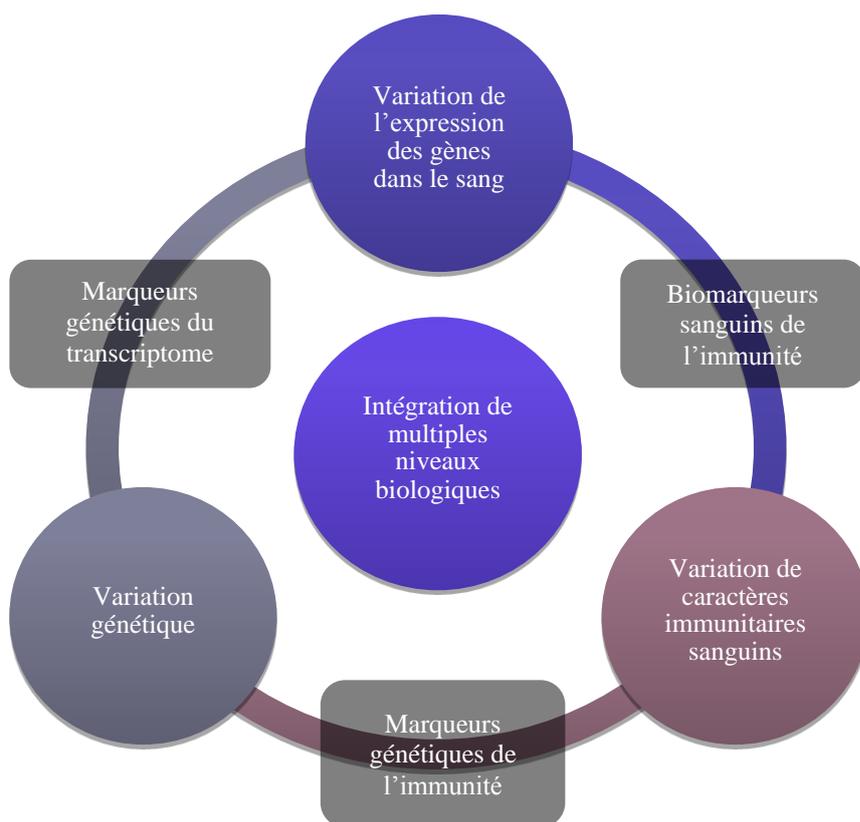


Figure O-23 : Représentation des étapes d'intégrations des trois niveaux biologiques abordés (cercles reliés) : phénotypique, transcriptomique et génétique. Les encadrés noirs donnent le résultat attendu pour l'intégration des deux niveaux auxquels ils sont liés.

Les objectifs définis pour ce travail de thèse sont les suivants :

1. Identification de marqueurs génétiques dans des variations d'expression des gènes dans le sang
2. Comparaison du transcriptome du sang et de tissus immunitaires spécialisés
3. Identification de biomarqueurs sanguins prédictifs de traits immunitaires
4. Identification de marqueurs génétiques pour la sélection sur des critères immunitaires

Les données disponibles pour remplir chaque objectif sont décrites dans la Figure O-24.

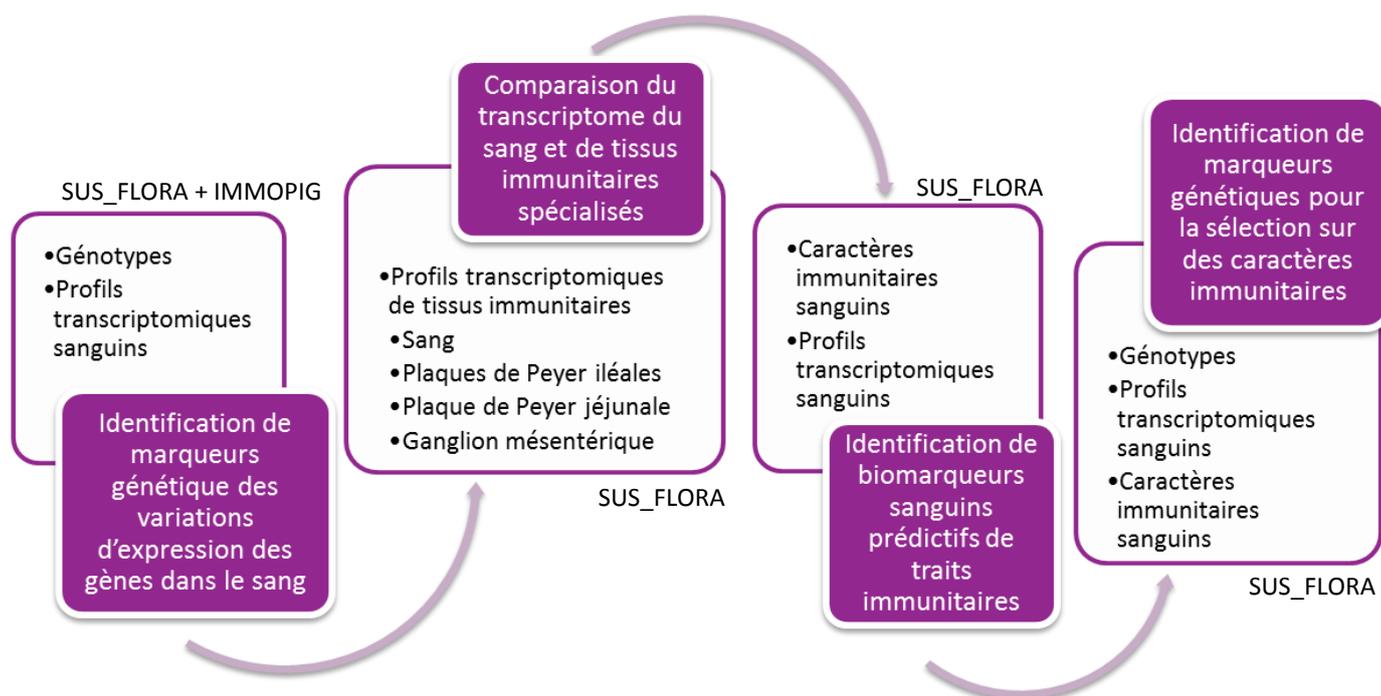


Figure O-24 : Représentation des objectifs spécifiques du travail de thèse dans les cadres à fond violet. Les données disponibles pour chaque objectif sont décrites dans les cadres à fond blanc. Les projets dont les données sont issues sont inscrits en dehors du cadre

I. Etude 1 : Cartographie du contrôle génétique du transcriptome sanguin

Article accepté à BMC Genomics le 25 novembre 2017

A. Objectifs et Méthodes

L'objectif de cette étude était d'établir une première cartographie du contrôle génétique du transcriptome du sang chez des porcs Large-White. Deux types d'analyses complémentaires ont été menées.

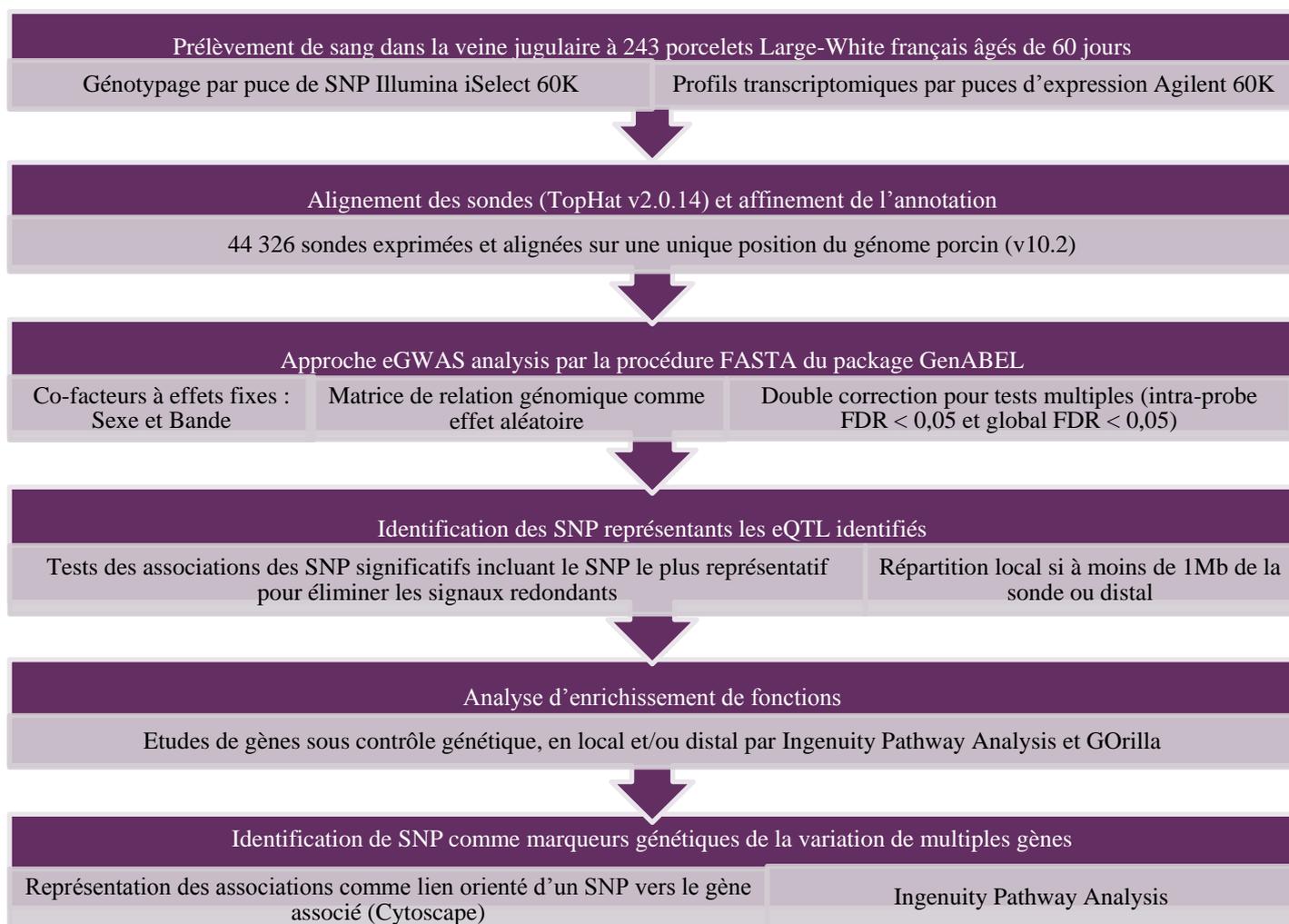


Figure E1-25 : Workflow d'analyse suivi pour l'eGWAS

La première partie de cette étude était une analyse d'association génome entier sur 242 animaux (SUS_FLORA) pour la variation d'expression dans le sang ou eGWAS (Figure E1-25) permettant l'intégration de profils transcriptomiques, définis par des puces d'hybridation Agilent-037880/INRA et des génotypes représentés par plus de 60 000 SNP étudiés par des puces de génotypage PorcineSNP60 DNA (Illumina).



La seconde analyse était une détection d'expression allèle spécifique (Figure E1-26) grâce à des données de RNA-Seq du sang périphérique pour 38 mâles non castrés (IMMOPIG).

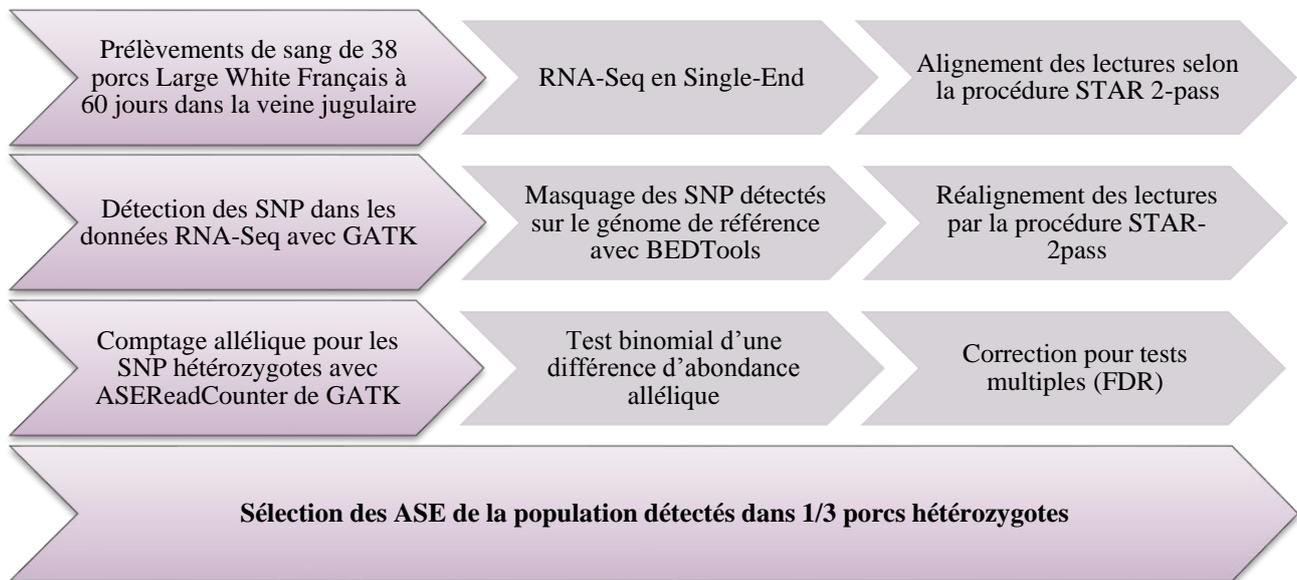


Figure E1-26 : Workflow d'analyse suivi pour l'analyse ASE

B. Résultats

L'approche eGWAS a mis en évidence 1 901 gènes associés à au moins un eQTL et 2 839 SNP-eQTL locaux et 1 752 SNP-eQTL distaux. L'analyse ASE a montré une expression différentielle pour 2 286 SNP dans les données de RNA-Seq. Ces SNP étaient localisés dans 763 gènes. La combinaison des deux analyses a mis en évidence que le contrôle génétique de l'expression des gènes est réparti de façon hétérogène sur le génome. Les résultats de l'approche eGWAS ont permis d'identifier des SNP-eQTL associés à l'expression de différents gènes constituant ainsi des réseaux à explorer. D'un point de vue fonctionnel, les gènes sous influence génétique sont principalement impliqués dans des voies biologiques liées à la maturation des ARN et à des fonctions immunitaires. Par ailleurs, les deux analyses ont révélé une forte influence génétique sur les gènes de la région du complexe majeur d'histocompatibilité (CMH) porcine (ou SLA pour *Swine Leucocyte Antigen*), située sur le chromosome 7. Plus précisément, de nombreux gènes du CMH de classe I et II contenaient des SNP avec des signaux d'expression allèle spécifique, laissant supposer un contrôle génétique *cis* de ces gènes, par un variant causal en déséquilibre de liaison.



C. Conclusion

Le transcriptome du sang périphérique chez le porc est soumis à un important contrôle génétique, avec plus de polymorphismes détectés en *cis* (local) qu'en *trans* (distant). Le nombre important de gènes immunitaires dont l'expression s'est révélée sous influence génétique dans le sang suggère que le transcriptome du sang est un phénotype intermédiaire intéressant, permettant de faire le lien entre variation génétique et variation phénotypique.



Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis

T. Maroilley^{1,*}, G. Lemonnier¹, J. Lecardonnel¹, D. Esquerré², Y. Ramayo-Caldas¹, M.J. Mercat³, C. Rogel-Gaillard^{1,*} and J. Estellé^{1,*}

¹ GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

² GenPhySE, INRA, INPT, ENVT, Université de Toulouse, 31326 Castanet-Tolosan, France.

³ IFIP - Institut du porc/BIOPORC, La Motte au Vicomte, BP 35104, 35651, Le Rheu, France.

*Corresponding authors

E-mail addresses:

TM: tatiana.maroilley@inra.fr

GL : gaetan.lemonnier@inra.fr

JL : jerome.lecardonnel@inra.fr

DE : diane.esquerre@inra.fr

YRC : yulixaxis.ramayo-caldas@inra.fr

MJM: marie-jose.mercat@ifip.asso.fr

CRG: claire.rogel-gaillard@inra.fr

JE: jordi.estelle@inra.fr

ABSTRACT

Background

Efforts to improve sustainability in livestock production systems have focused on two objectives: investigating the genetic control of immune function as it pertains to robustness and disease resistance, and finding predictive markers for use in breeding programs. In this context, the peripheral blood



transcriptome represents an important source of biological information about an individual's health and immunological status, and has been proposed for use as an intermediate phenotype to measure immune capacity. The objective of this work was to study the genetic architecture of variation in gene expressions in the blood of healthy young pigs using two approaches: an expression genome-wide association study (eGWAS) and allele-specific expression (ASE) analysis.

Results

The blood transcriptomes of 60-day-old Large White pigs were analyzed by expression microarrays for eGWAS (242 animals) and by RNA-Seq for ASE analysis (38 animals). Using eGWAS, the expression levels of 1,901 genes were found to be associated with expression quantitative trait loci (eQTLs). We recovered 2,839 local and 1,752 distant associations (SNP located less or more than 1 Mb from expression probe, respectively). ASE analyses confirmed the extensive *cis*-regulation of gene transcription in blood, and revealed allelic imbalance in 2,286 SNPs, which affected 763 genes. eQTLs and ASE-genes were widely distributed on all chromosomes. By analyzing mutually overlapping eGWAS results, we were able to describe putative regulatory networks, which were further refined using ASE data. At the functional level, genes with genetically controlled expression that were detected by eGWAS and/or ASE analyses were significantly enriched in biological processes related to RNA processing and immune function. Indeed, numerous distant and local regulatory relationships were detected within the major histocompatibility complex region on chromosome 7, revealing ASE for most class I and II genes.

Conclusions

This study represents, to the best of our knowledge, the first genome-wide map of the genetic control of gene expression in porcine peripheral blood. These results represent an interesting resource for the identification of genetic markers and blood biomarkers associated with variations in immunity traits in pigs, as well as any other complex traits for which blood is an appropriate surrogate tissue.



Keywords: blood, pig, transcriptome, eQTL, eGWAS, RNA-Seq, Allele-Specific Expression

Background

Modern livestock breeding programs aim to foster sustainability and improve production by selecting animals for health- and environment-related phenotypes. However, complex health traits such as immune capacity, disease resistance, and robustness are difficult to directly measure in living animals. There is a strong need for the identification of intermediate phenotypes that can be used as proxies for these traits. In pigs, data on immune function are regularly obtained by scoring innate and adaptive immunity traits measured in blood and serum (*e.g.*, blood cell counts, specific and non-specific antibodies, and acute phase proteins) or by studying the effect of *in vitro* stimulation of total blood or leucocytes (*e.g.*, production of chemokines and cytokines, phagocytosis) [1,2]. We and others have reported that many immunity traits are heritable in swine [1–3]. We have also shown that changes in the peripheral blood transcriptome are correlated to variation in immune traits [4]. In this context, then, peripheral blood appears to be a relevant tissue with which to phenotype immunity traits, as well as a relevant surrogate tissue for the quantification of other physiological traits.

Variation in transcript abundance has been studied as an inherited quantitative trait [5] and association mapping of loci for expression traits has been performed in several species for a wide range of tissues [6]. Indeed, as demonstrated by Schadt *et al.* [7], characterizing the relationships between DNA sequence and RNA expression is an appropriate step in understanding the links between genotype and phenotype. In the first global analysis of expression quantitative trait loci (eQTLs) conducted in mammals, these authors showed that gene expression traits in mice were connected with complex phenotypes like obesity. eQTLs can control the expression levels of local and/or distant transcripts [8] and are nowadays identified by expression genome-wide association studies (eGWASs). Local



and distant associations are often referred to as *cis*- and *trans*-associations and correspond to putative *cis*- and *trans*-regulatory relationships, which act, respectively, in an allele-specific and non-allele-specific manner. Complementarily to gene expression microarrays, an RNA-Seq-based approach provides sequencing data for the detection of both transcript expression levels and single-nucleotide polymorphisms (SNPs). By integrating these two types of data, allele-specific expression (ASE) analyses provide information on the relative expression levels of two alleles of a gene from the same individual, thus targeting *cis*-acting regulation [9]. An approach that combines eGWAS and ASE techniques is highly effective in characterizing candidate genetic variants acting as *cis*- and *trans*-regulators, as has been shown in the mouse for liver [10] and adipose tissue [11].

In humans, results on the genetic control of the blood transcriptome were reported to be robust and reproducible despite possible confounding effects due to variations in white and red cell counts [12]. These results highlighted the utility of blood eQTL mapping for improving the interpretability of GWAS results for complex phenotypes, especially by helping with the prioritization of candidate genes. Similarly, Schramm *et al.* [13] showed that eQTL analysis of whole blood is reliable and may be used to identify biomarkers and to enhance understanding of the molecular mechanisms underlying associations between genes and disease. Together, all these eQTL data have provided new types of resources and associated databases for the study of relationships between gene expression and phenotype of interest [12].

Few eGWASs have been reported in pigs. Liaubet *et al.* [13] analyzed gene expression in pig skeletal muscle sampled shortly after slaughtering, and showed an over-representation of genes that encoded proteins involved in processes induced during muscle development and metabolism, cell morphology, stress response, and apoptosis. More recently, eGWASs have been integrated with phenotyping studies in order to identify candidate genes and causative mutations associated with phenotypic variations in several tissues: for example, liver gene expression linked to blood and lipid traits [17],



gene expression in *longissimus dorsi* or *gluteus medius* muscles associated with growth, fatness, yield, and meat quality [14–19], and hypothalamus gene expression connected to coping behavior [20].

To our knowledge, no eGWAS of total blood has yet been performed in pigs. In this study, our aim was to build a genome-wide map of the genetic control of gene expression in the blood of 60-day-old Large White pigs by combining eGWAS and ASE results. Such a resource will be valuable for guiding integrated studies of complex phenotypes, as has been demonstrated in studies of humans [12], especially those linked to health and immune response in young pigs, which are prone to be subjected to infection challenges. After detecting the eQTL and ASE effects, we performed *in silico* functional analyses in order to illustrate the usefulness of our data in not only prioritizing candidate blood biomarkers and SNPs associated with immunity and physiological trait variations but also in deciphering the complexity of blood gene transcriptional regulation in pigs.

Results

eGWAS revealed numerous cases of local and distant regulation of gene transcription in blood

We performed an eGWAS by combining expression data from customized single-channel 8X60K Agilent arrays and SNP genotypes from Illumina PorcineSNP60K genotyping chips taken from 242 French Large White pigs. We searched for significant associations between all expressed probes that mapped to a unique position on the Sscrofa10.2 reference genome (39,649 probes from a total of 59,774 expressed probes) and all SNPs on the reference genome that met the quality control threshold (42,234 SNPs from a total of 61,557). For these association analyses, all probes, including those annotated to the same gene, were considered independent.

Overall, we identified 4,591 associations, which involved 3,195 eQTL-SNPs (see below) and 3,419 probes that were retained for further analysis (Figure 1 and Supplementary Table S1). Initially, 48,536 significant associations (FDR < 0.05) were detected between 18,541 SNPs and 3,419 expressed



probes mapping to a unique genomic position. In order to reduce redundancy in SNPs that was due to linkage disequilibrium, we represented the signals of all associated SNPs that harbored the same effect on probe expression by the most significant SNP in a window of 5 Mb. We used this SNP to represent the eQTL in further analyses and designated it as an eQTL-SNP. From this subset of eQTL-SNPs, 77% of the 3,419 associated probes could be assigned to 1,901 pig genes. Probe annotation revealed that gene assignment included alignments of sense (1,828 genes), antisense (73 genes), or both sense and antisense (244 genes) probe sequences (Supplementary Table S2). In addition, 421 probes could not be assigned to any known gene but were found associated with at least one eQTL-SNP.

A large majority of eQTL-SNPs (N=2,454) was associated with variation in the expression of a single probe. The other 965 eQTL-SNPs were each associated with between two (483 eQTL-SNPs) and 51 probes (one eQTL-SNP) (Table 1). Reciprocally, probes were associated with up to six eQTL-SNPs, although most were linked to only one or two (Table 1).

Among all associations, we characterized 2,839 as local and 1,752 as distant. In local associations, candidate eQTL-SNPs were located within 1 Mb of the start position of the associated probe and were expected to act on that probe's expression in an allele-specific manner, meaning a *cis*-regulatory action. All other associations were classified as distant associations and the corresponding eQTL-SNPs are referred to as distant eQTL-SNPs, expected to act in *trans* by regulating the associated gene in a non-allele-specific manner. Among the 3,195 eQTL-SNPs, 2,124 were local and 1,187 were distant, while 116 were involved in both local and distant associations. Furthermore, 2,407 probes were affected by at least one local eQTL-SNP and 1,427 probes by at least one distant eQTL-SNP, with 415 probes associated with at least one local and one distant eQTL-SNP. For 2,382 local associations (> 83%), the eQTL-SNP was located less than 500 kb from the start position of the probe, and for a subset of 1,334 associations, the eQTL-SNP was within 100 kb of the associated probe.



In order to evaluate the degree of the eQTL-SNP effect with respect to global variation in probe expression, we calculated the ratio of the eQTL-SNP effect to the standard deviation of the expression of its associated probe. The ratio distribution is represented in Figure 1A and showed that eQTL-SNPs had a wide range of additive effect sizes on the expression variation of associated probes. The means of local and distant eQTL-SNP effects were significantly different (unilateral Student test < 0.05), with local associations showing larger effects. Globally, the median ratio was equal to 0.8, meaning that for a majority of associations, the effect of the eQTL-SNP was nearly equal to the standard deviation of the gene probe expression. For 930 associations, the effect of the eQTL-SNP was larger than the standard deviation (ratio higher than 1; Figure 1A). In eight extreme cases, the effect was at least two times higher than the standard deviation of gene expression. The most extreme association had a ratio of 2.68; this involved the *IFITM2* gene on SSC2 and was also the most significant association found by eGWAS (q-value = 1.25×10^{-39}).

The analysis of eQTL genomic locations revealed that both local and distant associations were widely distributed on all chromosomes, although there were significant inter-chromosomal differences in the density of eGWAS signals (Figures 2A to 2D). Chromosomes SSC1, SSC8, SSC11, SSC13, SSC15, and SSC16 had significantly lower proportions of eQTL-SNPs than the genome as a whole, while SSC4, SSC7, SSC12, and SSC14 were relatively enriched in eQTL-SNPs (one-sided Fisher < 0.05). In addition, the relative proportions of local, distant, and local+distant eQTL-SNPs differed among chromosomes, especially on SSC1, SSC7, SSC12, and SSC15 ($\text{Chi}^2 < 0.05$). We also observed that the relative percentage of associated genes varied among chromosomes (Figure 2C). In particular, SSC1, SSC3, SSC6, and SSCX had lower percentages of associated genes than the genome as a whole (Fisher < 0.05), while SSC7, SSC10, and SSC14 were found to be enriched in associated genes (Fisher < 0.05). We also observed inter-chromosomal variation in the percentage of genes associated with local, distant, or local+distant eQTL-SNPs, as represented in Figure 2D. In particular, SSC6,



SSC12, SSC14, and SSCX presented a significantly different ratio between local- and distant-associated genes ($\text{Chi}^2 < 0.05$) in comparison to the proportion in the whole genome (Supplementary Figure 1). From these results, we targeted a subset of the most promising regions and networks, according to their level of significance and connectivity or their relevance to immune response, for further analyses.

ASE analysis confirmed extensive *cis*-regulation of gene expression in blood and revealed additional *cis*-regulatory events

ASE analysis was performed on whole-blood RNA-Seq data from a group of 38 Large White pigs. We found 12,811 Ensembl genes (release 82) expressed in blood; of these, 9,501 were assigned a HUGO gene symbol. To be considered expressed, a gene had to be represented by a minimum of 10 reads in at least one sample. Since genome re-sequencing of the 38 pigs was not available, SNP identification was carried out using the RNA-Seq data. Thus, this analysis only detected SNPs with different expressed alleles in the samples included in the ASE study, and was not able to detect silencing events.

After SNP calling on RNA-Seq reads, 82,419 SNPs met the quality control criteria, including 71,365 (86.2%) SNPs already known in the dbSNP database. Among these SNPs, 80,939 were heterozygous in at least one of the 38 samples, and 86.3% of these were already known in dbSNP. Overall, 6,824 genes were expressed and overlapped with at least one heterozygous SNP in one individual, and were thus detectable as affected by ASE.

From our analysis, 2,286 SNPs showed allelic imbalance (Supplementary Table S3). These SNPs were further referred to as ASE-SNPs. Of these, 63.5% had already been described in dbSNP (1,452 SNPs). They overlapped 1,312 transcripts and 989 Ensembl (release 82) annotated genes, of which 763 had a gene symbol and could be further explored with functional enrichment analysis. Overall,



these *cis*-regulated genes represented roughly 11% of all expressed genes that overlapped at least one heterozygous SNP. The genome localization of ASE-SNPs revealed that the candidate *cis*-regulated genes were widely distributed along chromosomes, with a great deal of inter-chromosomal variability. Specifically, enrichment in ASE-SNPs was detected on SSC5, SSC7, SSC9, SSC14, and SSCX, with the highest levels found on SSC7 (Figure 2E). In addition, SSC1, SSC11, and SSC13 were significantly depleted in genes affected by ASE-SNPs, while SSC3 and SSC5 were significantly enriched (Figure 2F, one-sided Fisher < 0.05). The distribution of ASE-SNPs along each chromosome was not homogeneous (Supplementary Figure S2).

In total, 149 genes were detected by both the eGWAS and ASE analysis (Supplementary Table S3); this corresponded to 7% of the associated genes found by eGWAS, 20% of genes affected by ASE, and 2.7% of all genes detected by both approaches. Among these 149 genes, 109 were represented by a sense-probe and associated with a local eQTL-SNP by eGWAS. This result provided validation of the genetic *cis*-regulation of these genes. Fifty-two genes that were affected by ASE were also found by eGWAS to be associated with a distant eQTL-SNP, and of these, 16 had a distant association with an eQTL-SNP on another chromosome. These 16 genes were thus candidates for being both *cis*- and *trans*-regulators.

Functional annotation of eQTL-SNPs, ASE-SNPs, and candidate regulated genes

The effects of eQTL-SNPs and ASE-SNPs were predicted using the Variant Effect Predictor (VEP) [21] and are summarized in Table 2. Of the eQTL-SNPs, 50.5% and 34% corresponded to intergenic and intronic variants, respectively. However, there were significantly fewer intergenic variants among eQTL-SNPs than in the SNP population as a whole (i.e. all SNPs that met QC criteria), while the opposite was true for intronic variants. With the ASE analysis, we observed that a majority of the detected SNPs with allelic imbalances corresponded to 3'UTR (24.5%), missense (23.6%), or



synonymous (18.1%) variants.

In order to explore whether the genetic control of gene expression in blood has an impact on specific biological pathways, we performed a functional enrichment analysis of the genes associated with local and/or distant eQTL-SNPs (Figure 3A) and genes that contained an ASE-SNP (Figure 3B). This analysis was carried out by first converting the pig gene lists into human gene lists using Biomart [22], in order to upload them into the GOrilla tool [23] to identify and visualize enriched gene ontology (GO) terms. For both eGWAS and ASE approaches, the top 20 most-enriched biological functions are presented in Figure 3, with the full lists for eQTL in Supplementary Tables S4 to S6 and for ASE in Supplementary Table S7.

Among the top 20 most-significant biological processes that were enriched among genes associated by eGWAS with local and/or distant eQTL-SNPs (Figure 3A), 12 were related to metabolic and catabolic processes and 3 to RNA processing (RNA and non-coding RNA). Other significantly enriched processes included the tri-carboxylic acid cycle, the cellular catabolic process, mitochondrial transport, and the organophosphate biosynthetic process. The two immunity-related functions on the list involved the interferon-gamma-mediated signaling pathway and the “presentation of exogenous peptide antigen via MHC class I, TAP-independent”, which was also among the top 20 enriched processes in the ASE analysis (Supplementary Table S7). In order to prevent bias in the enrichment analysis due to the inclusion of both local and distant associations, we also performed a separate enrichment study for each association type (Supplementary Tables S5 and S6, Figure 3A). There were 18 GO terms enriched among genes that were associated with distant eQTL-SNPs, while 7 terms appeared on the list generated by an analysis of genes linked to local eQTL-SNPs. The only function present in both lists was the ncRNA metabolic process. No immunity-related functions were shared by both local and distant associations; instead, the “interferon-gamma-mediated signaling pathway” was enriched in local-associated genes and the “antigen processing and presentation of exogenous



peptide antigen via MHC class I” function was identified only by combining all data. Three biological processes related to RNA processing were enriched among distant-associated genes. Overall, although a few processes were shared between genes associated with local or distant eQTL-SNPs, for the most part, the two groups of genes seemed to be involved in different biological pathways (Supplementary Tables S5 and S6 and Supplementary Figure S3). The high degree of overlap in enriched GO terms between the global analysis and the local-associated genes could be mostly due to the over-representation in the former analysis of genes associated with local eQTL-SNPs.

The genes that overlapped ASE-SNPs were significantly linked to a much higher number of biological functions than that identified from eGWAS data (Supplementary Table S7). In total, the ASE-genes corresponded to 105 enriched functions, with many of these related to immunity. Indeed, the top function listed among the 20 most-significant biological processes (Figure 3B) was the immune system process (GO:0002376), followed by 12 other processes associated with immunity and defense, response to stimulus, or viral transcription, and 5 processes associated with adhesion and signaling pathways, which included both type I interferon and interferon-gamma-mediated signaling pathways. In addition, *cis*-regulated genes revealed by the ASE analysis were also found to be enriched for processes related to translational initiation (GO:0006413), nuclear-transcribed mRNA catabolic process nonsense-mediated decay (GO:0000184), and regulation of translation (GO:0006417), which could all be linked to the genetic control of gene expression, especially by master regulators. Two functions linked to heme metabolism and biosynthesis were also significantly enriched among the *cis*-regulated ASE genes (GO:0006783: “Heme biosynthetic process”, GO:0006783: “Heme metabolic process”), which is unsurprising for genes expressed in blood. Only two GO functions were shared between the list based on ASE genes and that produced using the eGWAS approach (Supplementary Figure S3): the interferon-gamma-mediated signaling pathway (GO:0060333) and “antigen processing and presentation of exogenous peptide antigen via MHC class



I, TAP-independent” (GO:0002480).

Construction of association networks revealed that genes are functionally linked and under shared genetic control

Since several eQTL-SNPs were associated with variation in the expression of more than one gene, and many genes were also linked to more than one eQTL-SNP (Table 1), we were able to construct potential regulatory networks by overlapping the eQTL signals. The whole set of SNP-probe associations was visualized in Cytoscape v3.2.1 [24], which identified 831 independent networks involving more than two nodes (Table 3). A network of three nodes represented either two eQTL-SNPs associated with the same gene or two genes associated with the same eQTL-SNP. Most networks were simple (655 consisted of one eQTL-SNP affecting several genes). When all networks were taken together, nodes had on average 1.5 neighbors, and we found 329 multi-edge node pairs, which represented associations between a single eQTL-SNP and multiple probes assigned to the same gene.

The largest network (Figure 4A) was centered on eQTL-SNP rs81422644 (H3GA0029721), which mapped onto SSC10 at 30,039,892 bp (Supplementary Table S1). This eQTL-SNP was associated with the distant regulation of transcription in 51 probes, which were annotated to 47 distinct genes (three probes could not be annotated to any known porcine genes and for the *LNPEP* gene there were two probes). These 47 genes were widely distributed on all chromosomes, with the exception of SSC3 and SSCY (Supplementary Figure S4A). Using the R package PCIT [25], we applied the Partial Correlation and Information Theory algorithm [26] on expression correlations; this showed that variations in transcription levels of these genes were positively correlated regardless of SNP genotype (average correlation coefficient = 0.75; Supplementary Figure S4B). These analyses suggested that the 47 genes are genetically controlled by a common regulator in close vicinity to eQTL-SNP



rs81422644. However, only two individuals in our analyses were homozygous for the alternative allele of this eQTL-SNP, which means that this potentially interesting master regulatory SNP must be confirmed using another pig population in which this allele is present at a higher minimum allele frequency (MAF).

This large network was connected to another network that was centered around eQTL-SNP rs81393122 (ALGA0034786), which was located on SSC6 at 19,761,160 bp (Supplementary Table S1). This eQTL-SNP was linked to the distant regulation of 11 probes, annotated for ten genes on six distinct chromosomes (Figure 5A, Supplementary Table S1): ribosomal protein L3 (*RPL3*), ribosomal protein S20 (*RPS20*), cytochrome c oxidase subunit Vib polypeptide 1 (*COX6B1*), cytochrome c oxidase subunit VIa polypeptide 1 (*COX6A1*), non-metastatic cells 2 protein (*NME2*), taste receptor, type 1, member 3 (*TAS1R3*), integrin-linked kinase-associated serine/threonine phosphatase (*ILKAP*), glia maturation factor gamma (*GMFG*), uncharacterized protein C7H6orf226, and ecdysoneless cell cycle regulator (*ECD*). PCIT analysis revealed positive correlations among transcription variations in the 11 probes, with an average correlation coefficient of 0.61 (Figure 5B). We also observed that this eQTL-SNP genotype affected the transcription of the ten genes in the same direction (Figure 5C). Strikingly, nine of the ten associated genes were included in a unique IPA-predicted network related to the functions “cellular assembly and organization”, “cell death and survival”, “endocrine system disorders” (Figure 5D). One of the most significant associations in the network was with the gene *COX6B1* (Figure 5E).

As represented in Figures 4B and 4C, a candidate master regulatory SNP (rs81288717, or MARC0001946) was detected in an intergenic hotspot on SSC3, at 76,408,626 bp (Supplementary Table S1). This eQTL-SNP was associated with the distant-regulation of transcription variation in 26 genes, distributed on 14 different chromosomes (Supplementary Figure S5A). The eQTL-SNP genotype affected the transcription of the 26 genes in different directions, but with additive effects.



As an example, allele G was associated with increased transcription of the gene *UROCI*, but with decreased transcription of the gene *ARHGAP2* (Supplementary Figure S5B). Complementarily, PCIT analysis revealed that transcription variations in these 26 genes were strongly correlated, either positively (average correlation coefficient = 0.74) or negatively (average correlation coefficient = 0.47) (Supplementary Figure S5C). IPA highlighted that the 26 genes could be included in three different networks associated with immunity-related processes: ten genes were linked to “cellular development, hematological system development and function, hematopoiesis”, nine genes to “inflammatory response, cell signaling, molecular transport”, and five genes to “developmental disorder, hereditary disorder, organismal injury and abnormalities” (Supplementary Figure S5D). Thus, eQTL-SNP rs81288717 could be an interesting candidate genetic marker for the regulation of immunity-related genes in blood, but further work is needed to identify the causal genetic variant.

For a few networks, the central eQTL-SNP was located within 1 Mb of an ASE-SNP. As an example, we present the case of two eGWAS signals that were 295 Kb apart, both within 500 Kb of the same ASE-SNP (Figure 4C). The eQTL-SNP rs81338631 (ALGA0110383) maps onto SSC6 at 145,457,494 bp, inside an intron of *PCSK9*, while eQTL-SNP rs81244817 (MARC0056482) is located on the same chromosome at 145,752,528 bp (Supplementary Table S1). The nearby ASE-SNP, rs342721226, is contained within the 3'UTR region of the *YIPF1* gene, at 146,133,708 bp (Figure 4C). Association networks revealed that eQTL-SNP rs81338631 was linked to the distant-regulation of nine different genes located on seven distinct chromosomes, while eQTL-SNP rs81244817 was associated with three genes located on three different chromosomes (Supplementary Fig S6A). In a PCIT analysis, the transcription levels of genes associated with these two eQTL-SNPs were strongly positively correlated (0.63 on average; Supplementary Figure S6B). Moreover, IPA revealed that the whole set of genes associated with the two eQTL-SNPs, as well as the ASE-gene *YIPF1* and *PCSK9*, the gene that contained the eQTL-SNP rs81338631, could be included in a unique



functional network related to “cell death and survival”, “organismal injury and abnormalities” and “tissue morphology” (Supplementary Fig S8C). Both eQTL-SNPs affected the transcription of their associated genes in the same direction (Supplementary Figure S6D). Overall, these two regulatory networks seem to be related and the ASE affecting YIPF1 designed this gene as an interesting candidate for understanding the molecular mechanism of transcriptome regulation of associated genes (YIPF1 is associated to GO annotations that include RNA binding and ribonuclease activity).

A high density of local- and distant-acting eQTLs in the Swine Leucocyte Antigen (SLA) complex on SSC7

Chromosome SSC7 was enriched in both eQTL- and ASE-SNPs (Figure 2, Supplementary Figures S1 and S2). In particular, we observed numerous local and distant associations involving genes and SNPs within the MHC region, as has already been reported for human MHC [27–29].

The MHC in pigs is referred to as the SLA complex and maps to SSC7 on both sides of the centromere; the class I and III subregions cover 1.8 Mb of 7p1.1 band while the class II subregion spans 0.6 Mb of 7q1.1 band, with the size of the in-between centromere as-yet unknown [30]. Within this relatively small region, the SLA complex contains 151 annotated gene loci, and due to this high gene density and lack of recombination in this region, our previous definition of local/distant associations (inside/outside a 2-Mb window centered on the expressed probe) was not appropriate. Indeed, for the majority of eQTLs that mapped to the SLA complex, we could not distinguish between local and distant associations. Figure 6 illustrates the associations identified between gene expression levels and eQTL-SNPs, as well as genes affected by ASE-SNPs that have already been described in dbSNP. According to our eGWAS results, the expression levels of 51 genes within the SLA complex were found to be genetically controlled by 78 eQTL-SNPs, which included 11 SNPs located on chromosomes others than SSC7 and 31 SNPs on SSC7 but outside the SLA region. We identified 36



non-redundant eQTL-SNPs within the SLA complex and their VEP annotation is summarized in Figure 6. Of these 36 eQTL-SNPs, 29 mapped to annotated genes, which included four SLA genes (*SLA-7*, *SLA-11*, *SLA-DRB3*, and *SLA-DOA*) and 23 non-SLA genes. The remaining two were considered to be intergenic when only the most severe consequence of each SNP was considered in VEP analysis.

Twelve SLA genes and 39 non-SLA genes were found to be regulated by between one and five distinct eQTL-SNPs. Among those affected were four functional SLA class I genes, the classical class Ia genes *SLA-1* and *SLA-3*, as well as two non-classical class Ib genes, *SLA-7* and *SLA-8*. In the class II subregion, the expression levels of most genes were found to be genetically regulated, including *SLA-DRA*, *SLA-DRB1*, *SLA-DRB2*, *SLA-DQA*, *SLA-DOB2*, and *SLA-DOA*, as well as the non-SLA genes *TAP2* and *PSMB8*. All these data revealed that transcriptional activity is both dense and locally regulated. eQTL-SNPs were found to act either within a sub-region or more distantly, on genes mapping to both sides of the centromere. Only *SLA-DOA* and *TRIM26* were associated with an eQTL-SNP that had been annotated as a variant for these genes: a 3'UTR variant for *SLA-DOA* and a synonymous variant for *TRIM26*. These two genes were thus identified as good candidates for *cis*-regulation. We did not find eQTL-SNPs outside of *SSC7* that were associated with changes in the expression of SLA genes; the only associated eQTL-SNPs were located either within the SLA complex or in close vicinity. We detected 32 eQTL-SNPs specifically associated with a single gene (*SLA-1*, *SLA-3*, *SLA-5*, *SLA-8*, *SLA-11*, *SLA-DRA*, *SLA-DRB2*, or *SLA-DOA*). These SNPs could thus be considered genetic markers that specifically target the expression of one of these genes.

Through ASE analysis, we identified nine genes within the class I subregion and six genes within the class II subregion that harbored putative ASE-SNPs (Figure 6). Interestingly, each SLA gene was found to be affected by specific ASE-SNPs. *SLA-DQB1* and *SLA-DQB2* shared the same three ASE-SNPs because of their close vicinity. The number of ASE-SNPs per SLA gene varied. *SLA-3* harbored



84 SNPs with allelic imbalance, while *SLA-2* and *SLA-DRB1* showed 27 and 28, respectively. *SLA-DOA* and *SLA-DRA* harbored only 1 ASE-SNP. This observed variation in the number of SNP variants per SLA gene is consistent with the expected allelic variability according to each gene (www.ebi.ac.uk/ipd/mhc/group/SLA). The functional classification of ASE-SNPs for SLA genes is presented in Table 4.

Overlap of eQTL-SNPs and ASE-SNPs with blood biomarkers previously identified in pigs

We compared the lists of genes found using the ASE or eGWAS datasets to lists of genes that we previously reported to be differentially expressed (DE) in the blood of pigs; this included results from *in vitro* production of IL2 and IL10 after stimulation of total blood, *in vitro* phagocytosis tests, and counts of T lymphocytes CD4⁺/CD8⁺ [4]. For each immunity-related trait, numerous DE genes were associated with eQTLs or ASE (Table 5), indicating that expression levels of blood biomarkers for these traits are genetically controlled. For example, local eQTL-SNPs were associated with expression changes in the *ALOX12* gene, which has been identified as a blood biomarker for phagocytosis, as well as the genes *GZMB* or *KLRG1*, which are correlated CD4⁺/CD8⁺ cell count. Instead, two distant eQTL-SNPs were implicated in the transcription variation of *GATM*, which has been identified as a candidate biomarker for phagocytosis level; these eQTL-SNPs mapped less than 3 Mb from *GATM*. Finally, for CD4⁺/CD8⁺ cell count, the candidate blood biomarkers *TNLI*, *GNLY*, *GZMB*, and *KLRG1* were found to be affected by a SNP with allelic imbalance.

Discussion

eGWAS and ASE analysis provide complementary results for mapping genetic regulation of gene transcription in blood



In the current study we investigated the genetic control of the blood transcriptome in 60-day-old pigs by combining two high-throughput methods: i) an eGWAS which analyzed transcription in 242 pigs (130 uncastrated males and 112 females) using expression microarrays and genome-wide SNP genotyping, and ii) an evaluation of ASE in 38 castrated males based on RNA-Seq of total blood RNA. The groups of animals used for the eGWAS and ASE analysis were independent from each other but selected from the same larger experimental population of Large White pigs, and partly generated from common boars. In doing this, we aimed to prevent discrepancies in the results due to heterogeneity in genetic backgrounds. Each approach had a different focus: eGWAS maps eQTLs as genomic loci that influence levels of gene transcripts through local and/or distant actions [8], while ASE analysis identifies allelic imbalance among transcripts. In our study, eQTL-SNPs were defined as acting locally or distantly based on whether the SNP was less or more than 1 Mb from either side of the start position of the associated probe, respectively. Globally, we found 2,124 local eQTL-SNPs which affected the expression of 1,363 genes, and 1,187 distant eQTL-SNPs affecting 1,010 genes. In parallel, ASE analysis enabled us to detect 2,286 candidate SNPs that showed allelic imbalance, and which affected 763 genes.

Both eGWAS and ASE analyses were performed using stringent parameters to ensure robust mapping. For eGWAS, we chose to simultaneously analyze local and distant eQTLs [32–34], while previous studies had performed two independent association analyses which separated local- and distant-acting SNPs [27,29,35]. Due to the large amount of transcript information and the number of SNPs involved, eGWASs require extremely extensive calculations, which can raise issues related to statistical power [33]. To address this, we used stringent quality control thresholds, and corrected local and distant associations for multiple testing with an FDR procedure at two levels: i) for all association tests for each probe, and ii) for all association tests for all probes. For ASE analyses we filtered out SNPs with fewer than 10 reads and retained only ASE-SNPs with significant allelic imbalance in their read



abundances in at least one-third of heterozygous animals.

For both eGWAS and ASE approaches, we observed that the genetic control of gene transcription was heterogeneously distributed along the genome, with important variations among chromosomes in both numbers of eQTL- or ASE-SNPs as well as in the relative proportions of genes that demonstrated ASE or that were associated with eQTLs. However, we found only limited overlap between the local associations that were identified by eGWAS and the *cis*-associations that were identified by the ASE analyses, which is consistent with results from other studies [10,11]. Indeed, this was expected due to inherent differences between the approaches and their respective limits. For eGWAS, only eQTLs that corresponded to SNPs represented on the 60K SNP chip could be detected and in this design, intergenic SNPs are overrepresented [36]. Consequently, the functional annotation of eQTL-SNPs remains limited, and they mainly reflect cases of linkage disequilibrium with nearby causal variants and not direct causal links. The ASE analysis was limited by the lack of whole-genome sequence data for the animals included in this analysis, which meant that SNPs were called using RNA-Seq data. Therefore, in order to limit errors in SNP discovery, only ASE-SNP positions for which both alleles were expressed (Minor Allele Count > 3) were considered in the final analysis, and transcripts with low read counts were discarded. In addition, due to the lack of parental information, we could not differentiate between allelic imbalance and parental imprinting. However, among the 763 ASE genes found in our study, only 3 are described as imprinted genes in the Geneimprint database (<http://www.geneimprint.com/>): *SNRPN*, which is imprinted by the paternal allele in *Homo sapiens*, *Macaca mulatta*, and *Bos taurus*; and *SNORD64* and *DMTN*, which are reported as imprinted only in *Homo sapiens*.

Overall, all these limitations suggest that these eQTL-SNPs and ASE-SNPs are not causal variants but SNPs in linkage disequilibrium with the causal variants that could be used as genetic markers for gene expression in the pig population used in our study. We did not address the question of sex effect



on blood transcriptome variation but we included the sex as a cofactor in the eGWAS models and used only castrated males for the ASE analyses.

In a comparison of ASE and eGWAS results, 149 associated genes in eGWAS overlapped with ASE-SNPs. For genes found *cis*-associated in eGWAS, it validated the putative *cis*-regulation of their expression. For other genes, ASE analysis allows refining eQTL-SNPs annotation as *cis* when they mapped at a few more than the arbitrary distance of 1 Mb. By combining eGWAS and ASE analyses, we were able to more extensively map the genetic control of the blood transcriptome and this approach will likely contribute to a more fine-scale mapping of the candidate causal polymorphisms. The heterogeneity of associated SNPs along the genome can be observed with greater precision using ASE approaches than with eGWAS, and this precision enabled us to pinpoint the most relevant genomic regions, as illustrated by the example of the SLA region. At the functional level, we found that ASE-genes were more enriched in biological functions specifically related to immunity than the genes locally or distantly associated with eQTL-SNPs were. However, the few enriched GO terms that were shared between the two analyses were immune-related. The GO results were complemented by the fact that the ASE analysis pinpointed immune functions as one of particular phenotypes affected by gene expression found under genetic control. These findings suggest that ASE results are more precise and accurate than eQTL results. In our study, this difference could be related to the relatively low density of SNPs in the chip used for eQTL mapping. Finally, ASE can provide information relevant in refining the mechanism of genetic regulation of gene expression by completing for instance putative regulatory networks.

Results are consistent with previous eQTL studies of human whole blood

Globally, our results provided general patterns of eQTL mapping in blood that were consistent with previous results on whole blood in humans. First, we detected 1.6 times more local associations (2,839)



than distant associations (1,751), an imbalance that has already been reported in previous studies [12,32,34,37]. This could be because distantly acting regulators have weaker effects than locally acting ones, resulting in associations that require greater statistical power for detection. Moreover, *trans*-eQTLs are more cell specific than *cis*-eQTLs [29] and because the heterogeneous nature of blood, *trans*-signals will be weaker. Although we cannot preclude the explanation that the larger number of local associations recovered here is due to a lack of statistical power to detect distant/*trans*-regulation, our results suggest that the number of local/*cis*-regulators truly does exceed the number of distant/*trans*-regulatory SNPs.

We observed that local eQTLs, distant eQTLs, and *cis*-regulators were widely distributed across all chromosomes, as has been previously described [32,34]. Interestingly, nine genes that were associated with local eQTL-SNPs, two of which belonged to the *TMEM* family, had also been listed among the top 25 *cis*-associated genes reported by Joehanes *et al.* [12]. Likewise, a set of 324 *cis*-acting genes reported by Mehta *et al.* [34] in a study on whole blood also contained 32 of the *cis*-associated genes found here. Moreover, as has also been reported in humans, we were able to identify putative candidate master regulatory variants, despite the heterogeneous nature of peripheral blood [28,29,32]. We detected at least 13 master regulatory loci, as represented by eQTL-SNPs that were associated with between 10 and 51 probes. These eQTL-SNPs influenced the expression of various genes, mostly in the same direction, as illustrated here by the examples of two SNPs, ASGA0110383 and MARC0056482. However, most master regulators were not associated with local genes, which was consistent with the results of a previous study [32].

We paid special attention to the SLA region on chromosome SSC7 because it harbored the greatest concentration of genetic regulatory elements in both eGWAS and ASE analyses. Previous studies in humans have reported similar findings. For example, Fairfax *et al.* [29] found that several loci in the MHC showed significant *trans*-associations. Examination of the MHC also revealed numerous



eQTL-SNPs linked to complex diseases [27]. In addition, an eGWAS performed on blood from cohorts of healthy women and breast cancer survivors showed an association between ten HLA genes (*HLA-C*, *HLA-E*, *HLA-F*, *HLA-G*, *HLA-H*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB3*, *HLA-DRB4*) and SNPs in 100 genes located on human chromosome 6 [38]. In particular, *CIITA* has been identified as a master *trans*-activator, with an essential role in initiating transcription of MHC class II genes in human stimulated B cells and monocytes [39]. However, our eGWAS did not generate any evidence of *CIITA* acting as a master regulator, despite the presence of one *CIITA*-related SNP on the array. In the ASE analysis, *CIITA* was expressed in the RNA-Seq data, but had no significant allelic imbalance. A possible explanation for this discrepancy is that such associations could be specific to conditions of immune stimulation and are not significant in a basal health state. Strikingly, we observed that almost all functional SLA class I and II genes harbored allelic imbalance, suggesting that fine-scale modulations of the expression levels of histocompatibility class I and II molecules may depend on the relative abundance of alleles. More in-depth analyses are required to determine whether this allelic imbalance affects peptide presentation efficiency and is associated with different immunity phenotypes.

Our results seem to be globally consistent with studies on the whole blood transcriptome in humans [28,29,32]. However, additional studies are needed on whole blood in pigs to assess the robustness of the list of local eQTL-SNPs reported here and to start to establish comparative eQTL maps where relevant. Thus, our work is a first step in generating useful resources for the study of eQTLs in pigs, as has been established in humans [12].

A step forward in the characterization of blood biomarkers, particularly for immune capacity

Peripheral blood contains various types of important immune cells. Schramm *et al.* [32] demonstrated reasonable concordance between the *cis*-eQTLs found in their own study of whole blood and those



reported in prior studies of primary monocytes or blood-derived lymphoblastoid cell lines. They thus confirmed that whole-blood eQTL studies are a good resource for the discovery of biomarkers, especially in the context of disease in humans. With the goal of improving selection for health traits in pigs, blood is also a relevant surrogate tissue for phenotyping immune capacity [40]. As shown in human studies, whole blood can be used for robust eQTL analysis despite its heterogeneous cell composition. Indeed, Mehta *et al.* [34] demonstrated that variations in cell count would exert only a minor effect on expression levels, as eQTLs are consistently replicated between whole blood and independent human cohorts.

We showed a notable degree of overlap between the genes found in this study to be genetically controlled and the QTLs available in the AnimalQTL database [31] for pigs. More precisely, we found overlap between the lists of eQTL- and ASE-SNPs generated here and QTLs of the AnimalQTL database that had been associated with immune-related phenotypes, including interferon-gamma level, IgG level, lymphocyte count and other traits related to blood function (Supplementary Table S8 and Table S9). In addition, we identified the genetic regulation of several previously published blood biomarkers for immunity traits (e.g., *ALOX12* for phagocytosis, *GZMB* and *KLRG1* for CD4-/CD8+ cell count) [4], which suggests that these biomarkers could be used as molecular expression phenotypes in genetic selection efforts.

The integration of eGWAS- and ASE-based approaches, together with GWAS data when relevant, facilitates the identification of candidate genes and polymorphisms that are related to complex traits. Similar integrated approaches in pigs have been recently published for complex traits related to meat properties [15,17], as well as for blood lipid traits linked with cardiovascular diseases [41] and traits associated with coping behavior [20]. In addition, this approach could help to pinpoint the biological and molecular bases of phenotype-genotype links that are highlighted by a GWAS-based approach. Integrated methods are also expected to be more accurate in the detection of loci under genetic control



as the link between genotype and transcriptome is more direct than the link between genotype and end phenotype. These techniques can thus help to prioritize the study of candidate variants linked to changes in the transcriptome and phenotype.

Conclusions

This study is, to the best of our knowledge, the first genome-wide analysis of the genetic control of gene expression in blood in the pig, based on a cohort of 60-day-old French Large White pigs. The combination of eGWAS and ASE results showed extensive genetic regulation of gene transcription, and provides a relevant resource for the study of genomic regions and variants linked to gene expression variation in pig blood. All results are available in a public Track Hub entry that can be explored using the genome browsers of Ensembl and UCSC. Additional studies are needed to cover more ages, breeds and environments in order to build an even more extensive map of genetic regulations of gene expression in blood. Overall, these data and approaches should contribute to future efforts to decipher the ultimate molecular mechanisms behind the genetic determination of pig phenotypes, especially for immunity- and health-related traits.

Methods

Animals and blood sampling

Two groups of pigs from the same French Large White selected line were bred either on a test farm at Le Rheu (38 pigs, IMMOPIG project) or on an INRA experimental farm at le Magneraud (243 pigs, SUS_FLORA project). The animals were partly generated from common boars and were weaned at 28 days of age. The group of 38 pigs belongs to a population of castrated males that was described in Flori *et al.* (2011) [2]. The group of 243 pigs comprised 131 uncastrated males and 112 females from a larger cohort that is described in Ramayo-Caldas *et al.* (2016) [42]. Details on the pig feed diet are available in Mach *et al.* (2015) [43] and correspond to standard nutritional practices in pig production.



Peripheral blood (jugular vein) was sampled from 60-day-old pigs using EDTA-coated tubes and PAXgene Blood RNA tubes (PreAnalytiX, Qiagen, Germany). Blood samples were stored at -20°C (EDTA tubes) or -80°C (PAXGene tubes) prior to DNA or RNA extraction procedures, respectively. Complete blood cell counts were recorded for a subset of 195 pigs from the group of 243 pigs (Supplementary Table S10), in agreement with our observations that the animals showed no clinical signs of infection. Information on production traits are available in Sanchez *et al.*, (2014) [44] for the ASE pig group and in Ramayo-Caldas *et al.* (2016) [42] for the eQTL pig group.

Total RNA extraction from blood

Total RNA from blood was isolated using the PAXgene Blood RNA Kit (Qiagen, Germany) following the manufacturer's instructions. RNA purity and concentration were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA), and RNA integrity was assessed using the Bioanalyzer 2100 (Agilent Technologies, USA). RNA samples used for RNA-Seq had RINs ranging from 6.7 to 8.6 with 28S:18S ratios ranging from 0.8 to 1.6. RNA samples used for microarrays had RINs between 6.1 and 9.7.

SNP Genotyping

DNA was extracted from blood samples and genotyped using the Illumina PorcineSNP60 DNA chip which contains over 60,000 SNPs [36]. Genomic DNA extraction and SNP genotyping were carried out with the Labogena-DNA platform (Jouy-en-Josas, France). The initial quality control (QC) step of genotyping was performed using the R package GenABEL [45] and included heterozygosity testing and multidimensional scaling to identify population outliers. For SNP filtering, we used sample and SNP call rates > 95%, a MAF > 5%, a Hardy-Weinberg equilibrium threshold corrected by FDR of 1%, and a XXY call of 0.8. We removed one individual that did not meet these QC criteria, resulting in 242 individuals (130 males and 112 females) with data available for 44,281 SNPs.



Analysis of blood transcriptome using expression microarrays

Total blood RNA (200 ng) was retro-transcribed, Cy3-labeled, and hybridized onto customized single-channel 8X60K Agilent Technologies arrays (platform Agilent-037880/INRA *Susscrofa60K* v1), which were enriched in genes linked to immunity and muscle physiology, following the manufacturer's instructions and a protocol described in Jacquier *et al.* (2015) [46]. Because genomic sequencing data were not available, it was not possible to take a possible effect of within cDNA target polymorphisms on hybridization levels into account. Pre-processing and quality control were performed by the Limma R package [47] on 60,306 probes. In brief, we first normalized the background by the normexp method. Then, intensities quantified by median signals were log₂-transformed and centered. Expression values were normalized by the quantile method. Probes were filtered according to their intensities in order to eliminate unexpressed genes; to be retained, probes had to be expressed at levels at least 10% higher than those of the negative controls on at least four arrays. In total, 59,774 probes that passed the quality control and the filtering step were considered to be expressed in the blood samples included in this study.

Refinement of probe sequence annotation of the Agilent-037880/INRA *Susscrofa60K* v1 platform

Probe mapping and annotation were updated to refine all genomic locations. The 60-mer sequence probes included in the Agilent-037880/INRA *Susscrofa60K* v1 platform were mapped onto the pig reference genome (SusScrofa v10.2) with the TopHat aligner (v2.0.14)[48]. Allowing two mismatches and two gaps per probe sequence, we found 44,326 probes mapping to unique genomic positions. Among these, 41,839 probes were expressed in our samples. The annotation of these probes was performed by combining annotation information provided by the Ensembl (release 79) and NCBI (GCF_000003025.5) databases.



Expression genome-wide association study (eGWAS)

The eGWAS was carried out on data from 242 pigs and linked SNP genotypes, based on animal genotypes obtained with PorcineSNP60 DNA chips, to blood gene-transcript levels, as detected by transcriptome analysis using Agilent-037880/INRA *Susscrofa*60K v1 microarrays. The association studies were performed with the GenABEL package [45] using the family-based score test for association [49], a polygenic mixed model to estimate the SNP effect, and a p-value in two steps. First, the polygenic additive model and likelihood were estimated using available data, and then the FASTA test statistic was computed using the maximum-likelihood estimates of the intercept, the proportion of variance explained by the polygenic component, and the residual variance. The model was corrected for different co-factors, with batch and sex as fixed effects, and with the familial structure of the sample as a random effect, represented by the genomic kinship matrix estimated using autosomal data. The individual variability in blood cell counts was not included in the analysis model since the data were not available for all the animals. Nonetheless, we observed a limited variability of these blood parameters within the population (see Supplementary Table S10). All probes and SNPs were treated independently for the first step of the association study. Next, multiple testing corrections were conducted at two levels. First, for each probe, p-values of the associations between the expression variation of that probe and all SNPs were corrected by an FDR procedure that controlled the overall type-I error rate at 5%. Secondly, to assess the significance of associations of all probes, a second significance threshold was calculated by FDR (5%) for the p-values of all associations. Only associations with a FASTA p-value lower than these two thresholds were considered significant. Finally, in order to facilitate the global interpretation of results, we implemented additional association tests to limit redundancy. We reduced the definition of an eQTL to a representative SNP, i.e. for each eQTL region, we kept only the most significant associated SNP to represent the eQTL, referred to hereafter as an eQTL-SNP. In order to eliminate other significant associated SNPs in a 5-Mb window around a designated eQTL-SNP without missing an independent eQTL, we repeated the



association analysis between the probe and these SNPs, correcting the polygenic mixed model by the effect of the most significant SNP, as has been reported in previous studies of human data [29,50]. If the p-values associated with a given SNP did not reach the previously calculated significance threshold, the SNP was considered to have the same effect as the most significant SNP (the previously designated eQTL-SNP); it was thus considered redundant and removed from further analyses.

Allele-specific expression (ASE) analysis by RNA-Seq

Sequencing libraries from total blood RNA (1.5 µg) were prepared using the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, USA) according to the manufacturer's instructions at INRA's GetPlage platform (Toulouse, France) and as previously reported in Mach *et al.* (2014) [51]. Tagged cDNA libraries were sequenced on 2.66 lanes of a HiSeq2000 (Illumina) in 100-bp single-end reads. Then, raw reads were aligned to the pig reference genome (v10.2) with the STAR 2.4.0 aligner using an approach that comprised two alignment steps, as recommended by the Broad Institute (<https://software.broadinstitute.org/gatk/guide/article?id=3891>). First, splice junctions were detected by a first pass of STAR mapping. Detected splice junctions were then used during a second round of mapping to take into account the discontinuous nature of the RNA sequencing data, allowing only one mapping position per read. STAR includes a soft-clipping step, which takes care of adapter contamination. After the two-pass procedure, we used a GATK (v3.4, [52]) tool called SplitNCigarReads, which split reads into exon segments and hard-clips any sequences overhanging into the intronic regions to reduce the number of called false variants. Then, reads were realigned around indels to limit artifacts due to the alignment algorithm. Finally, the standard GATK base recalibration procedure re-adjusted the quality scores assigned to the individual base calls in each sequence read, thus reducing the systematic technical error linked to the sequencing machines.

In order to prevent false positive ASE signals due to allelic mapping biases that could influence allele counts, variants were called in the dataset, filtered using vcftools (v.0.1.12a), and then nucleotides



that were characterized as SNPs were masked in the reference genome with maskFastaFromBed (Bedtools v2-2.24.0). Reads were then realigned on the masked reference genome using the same procedure as described previously in order to have the same constraint of one mismatch for reads that had either of the two alleles at the SNP position.

Allele counts were performed using a GATK (v.3.4) tool called ASEReadCounter which calculates allele counts at SNP positions. A binomial test determined whether there was a significant difference in abundance of each allele of a SNP. We corrected p-values on a per-animal basis using the FDR procedure ($5 \cdot 10^{-2}$). Only bi-allelic SNPs with a significant differential abundance in alleles in more than 1/3 of heterozygous animals for the considered SNP were analyzed further.

Analysis of function enrichment and SNP annotation

A functional enrichment analysis of genes associated with eQTL- or ASE-SNPs was performed with GOrilla (Gene Ontology enRIchment anaLysis and visualizAtion tool, Eden *et al.*, 2009). The Gorilla database is periodically updated using the GO database (Gene Ontology Consortium) and other sources and we used the updated version of December 2016. The p-value threshold was equal to $1 \cdot 10^{-3}$ and an FDR correction was performed on p-values. For gene networks, we also performed functional analyses that used QIAGEN's Ingenuity Pathway Analysis (IPA®, QIAGEN, Redwood City, CA, USA, <http://www.qiagen.com/ingenuity>, release date 2016-12-05) with default parameters. eQTL- and ASE-SNPs were annotated using the Variant Effect Predictor [21] to determine their known effects on gene, transcript, and protein sequences.

Visualization of results

To visualize eGWAS results, we used Cytoscape v3.2.1 [24], which enables the construction of association networks. We used genes and eQTL-SNPs as nodes and significant associations as edges. The R package RCirco [53] was used to illustrate master regulators, which we defined as SNPs that had associations with multiple genes. To generate Venn diagrams, we used jvenn, a plug-in for the jQuery Javascript library [54]. Boxplots were produced with the R package ggplot. The R package



PCIT [25] was used to analyze the correlation between the expression variation of probes associated with the same eQTL-SNP.

List of abbreviations

ASE: Allele-Specific Expression

eQTL : expression Quantitative Trait Loci

eGWAS : expression Genome-Wide Association Study

FASTA : Family bAsed Score Test for Association

GO : Gene Ontology

GWAS : Genome-Wide Association Study

IPA: Ingenuity Pathway Analysis

MHC : Major Histological Complex

PCIT: Partial Correlation Information Theory

QC: Quality Control

QTL: Quantitative Trait Loci

SNP : Single-Nucleotide Polymorphism

SLA : Swine Leucocyte Antigen

VEP : Variant Effect Predictor

Declarations

Ethics approval and consent for participation

All animal procedures were conducted according to the guidelines for the care and use of experimental animals established by INRA (Ability for animal experimentation: A78-172; agreement for experimentation at Le Magneraud: A-17661; protocol approved by the COMETHEA



ethics committee at Poitou-Charentes with permit number CE2013-2).

Consent for publication

All authors have approved the manuscript.

Availability of data and materials

Expression microarrays: All microarray data are MIAME compliant and have been submitted to NCBI's Gene Expression Omnibus with accession number: GSE97374.

RNA-seq data: The raw reads are available at NCBI's SRA repository (Bioproject PRJNA360366; accessions SRR5155448 to SRR5155485).

Track hub: All significant eGWAS and ASE results are available in a public Track Hub entry (https://www.trackhubregistry.org/search/view_trackhub/AVszgl3NpL2YcDCmVC8X) that can be explored using Ensembl's and UCSC's genome browsers.

Competing interests

The authors declare they have no competing interests.

Funding

TM was financially supported by a PhD fellowship co-funded by the INRA Animal Genetics Division and BIOPORC (association composed of AXIOM, CHOICE GENETICS FRANCE, NUCLEUS, and IFIP). Experiments were funded by the Sus_Flora (ANR-10-GENM-016) and IMMOPIG (ANR-06-GANI-08) projects of the French National Agency and INRA's Animal Genetics Division.

Authors' contributions

CRG coordinated the Sus_Flora and IMMOPIG projects. JE, TM, and CRG designed the experiments and interpreted all results. TM performed all analyses and was supervised by JE for computational analyses. YRC contributed to network analyses. CRG, JE, and GL contributed to animal sampling; GL organized all procedures for animal sampling and safe sample storage. GL



and JL performed RNA extraction and microarray hybridization on the @BRIDGE platform (<http://abridge.inra.fr>). DE performed RNA-Seq on the Genotoul platform (<http://bioinfo.genotoul.fr>). TM drafted the manuscript which was further completed by CRG and JE. All authors have read and improved the manuscript.

Acknowledgments

We are grateful to the Genotoul bioinformatics platform at INRA de Toulouse Midi-Pyrenees (BioinfoGenotoul) for providing computing and storage resources. We are grateful to the INRA MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for help and support. We are grateful to Laurence Liaubet (INRA, France) for sharing information on the expression microarrays and to Maria Ballester (IRTA, Spain) for useful comments and suggestions during the data analysis. We are grateful to the INRA @bridge platform for the secure storage of samples and the personnel at GENESI's farm at Le Magneraud for their implication on the generation of animals and samples.



Tables

Table 1. Numbers of significant associations between probes and eQTL-SNPs, and vice versa

	Number of associated eQTL-SNPs or probes											Total	
	1	2	3	4	5	6	7	8	9	10	11-51		
Probes	2423	839	141	14	1	1							3419
eQTL-SNPs	2454	483	146	44	23	16	7	7	2	4	9		3195

Table 2. Summary of the effects of SNPs identified by eGWAS and ASE, and their relative proportions in the larger population of a given SNP category (%).

SNP effects	eGWAS			ASE
	QC-SNPs ¹	eQTL-SNPs	P-values ²	ASE-SNPs
3'UTR variant	274 (0.65%)	47 (1.47%)	1.71x10 ^{-6*}	560 (24.5%)
5'UTR variant	51 (0.12%)	5 (0.16%)	5.95x10 ⁻¹	51 (2.2%)
Downstream gene variant	1278 (3%)	150 (4.7%)	9.03x10 ^{-7*}	125 (5.5%)
Intergenic variant	27211 (64.7%)	1606 (50.5%)	4x10 ^{-56*}	150 (6.6%)
Intronic variant	10893 (25.9%)	1083 (34%)	1.4x10 ^{-22*}	209 (9.1%)
Missense variant	187 (0.44%)	28 (0.87%)	1.79x10 ^{-3*}	539 (23.6%)
Non-coding exon variant	18 (0.04%)	3 (0.09%)	1.80x10 ⁻¹	127 (5.6%)
Splice region variant	68 (0.16%)	7 (0.21%)	3.70x10 ⁻¹	14 (0.6%)
Stop lost	7 (0.016%)	1 (0.03%)	4.42x10 ⁻¹	5 (0.2%)
Stop retained variant	16 (0.038%)	4 (0.13%)	4.78x10 ^{-2*}	1 (0.04%)
Synonymous variant	553 (1.31%)	68 (1.19%)	3.45x10 ^{-4*}	414 (18.1%)



Upstream gene variant	1526 (3.63%)	182 (5.7%)	1.92x10 ^{-8*}	51 (2.2%)
Mature miRNA variant	1 (0.002%)	0		1 (0.04%)
Splice acceptor variant	1 (0.002%)	0		10 (0.4%)
Stop gained	5 (0.012%)	0		17 (0.7%)
Splice donor variant				12 (0.5%)

¹ SNPs that met Quality Control criteria

² P-values for a Fisher test. For P < 0.05, the proportion of the SNP effect is significantly different between that of SNPs that passed the QC and that of eQTL-SNPs. Significant differences are labeled by a *.

Table 3. Number and size of putative regulatory networks constructed from eGWAS-based associations

Number of nodes per network	3	4	5	6	7	8 - 154	Total
Number of networks	488	166	73	42	13	41	831

Table 4. Number and annotation of ASE-SNPs affecting MHC genes

	SLA-1	SLA-2	SLA-3	SLA-5	SLA-7	SLA-11	SLA-DRA	SLA-DRB1	SLA-DOA	SLA-DQB1/2
3'UTR variant	13	15				6	1			3
5'UTR variant						1				
Intron variant	17	5	13		2			2		
Missense variant	28	2			1	7		14		
Non-coding transcript exon variant	2		61	12						
Splice region variant	12	1	1							
Synonymous variant						3		6	1	



Downstream gene variant	4	12		3	3		6			
Intergenic variant										
Total*	72	27	84	12	3	20	1	28	1	3

* Total of ASE-SNPs per gene identified in the 38 pigs included in the analysis. The total numbers do not correspond to the sum of annotated SNPs because a few SNPs were annotated differently depending on the transcript they overlapped.

Table 5: Correspondence between sets of eQTL- and ASE-SNPs detected in this study and gene sets that were differentially expressed (DE) in the blood of pigs studied for levels of immunity traits.

	IL2¹	IL10²	Phagocytosis³	CD4-/CD8+⁴
DE genes [3]	850	733	1195	52
DE local-associated genes	69	64	107	5
DE distant-associated genes	56	47	74	0
DE genes with ASE signal	48	29	61	7

¹ Production of IL2 after *in vitro* stimulation of total blood with phorbolmyristate acetate (PMA) and ionomycin for 48h

² Production of IL10 after *in vitro* stimulation of total blood with PMA and ionomycin for 48h

³ Quantification of phagocytosis capacity via *in vitro* tests on total blood

⁴ CD4-/CD8+ lymphocyte count measured by flow cytometry



Figure legends

Figure 1: Overview of associations identified by eGWAS

A: The distribution of additive effects of local-acting (red bars) or distant-acting (blue bars) eQTL-SNPs. eQTL-SNP effects are expressed as the ratio between the effect of the eQTL-SNP on probe expression and the standard deviation of variation in probe expression. B: Genomic locations of expressed probes versus genomic positions of associated eQTL-SNPs. Each significant association is represented by a dot; the color of dots indicates the significance (q-value) of the association (green: $1.10^{-1} < q < 1.10^{-10}$, dark red: $1.10^{-10} < q < 1.10^{-20}$, blue: $1.10^{-20} < q < 1.10^{-30}$, black: $1.10^{-30} < q < 1.10^{-40}$). Dots on the diagonal represent local associations; off-diagonal dots correspond to distant associations.

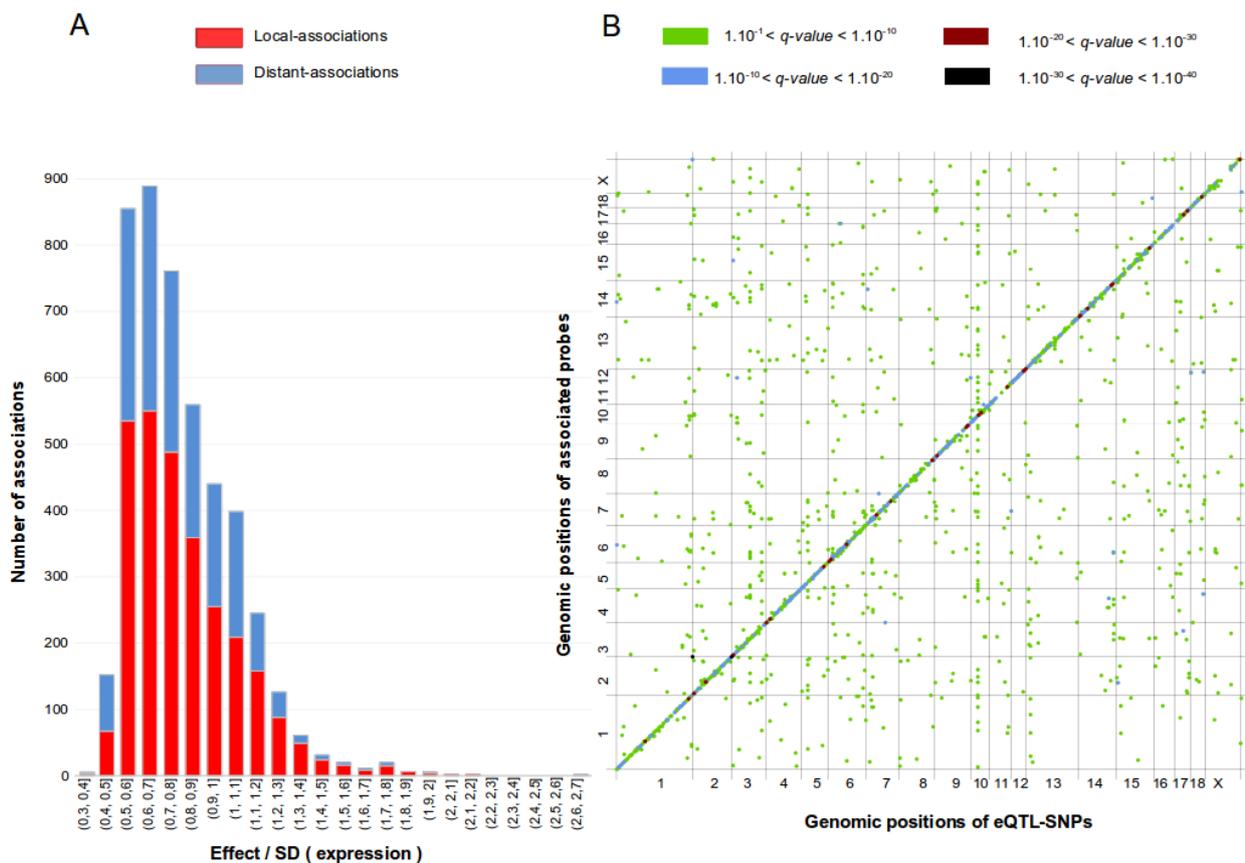


Figure 1



Figure 2: Genome-wide distribution of associations identified by eGWAS (A-D), and ASE-SNPs or -genes (E-F).

A: Per-chromosome percentage of eQTL-SNPs detected by eGWAS (red) in the larger SNP population. B: Relative proportions of SNPs that were detected as local-acting (red), distant-acting (blue) or both local- and distant-acting (purple) eQTL-SNPs by eGWAS. C: Per-chromosome percentage of genes whose transcription levels were linked to eQTL-SNPs. This percentage was calculated from the whole set of genes detected as expressed in blood. D: Percentages of genes associated with local-acting (red), distant-acting (blue) or both local- and distant-acting (purple) eQTL-SNPs. E: Per-chromosome percentage of ASE-SNPs of the whole set of 80,939 heterozygous SNPs in the group of 38 pigs included in the ASE analysis. F: Per-chromosome percentage of genes harboring at least one ASE-SNP in the set of genes with heterozygous SNPs in the group of 38 pigs.

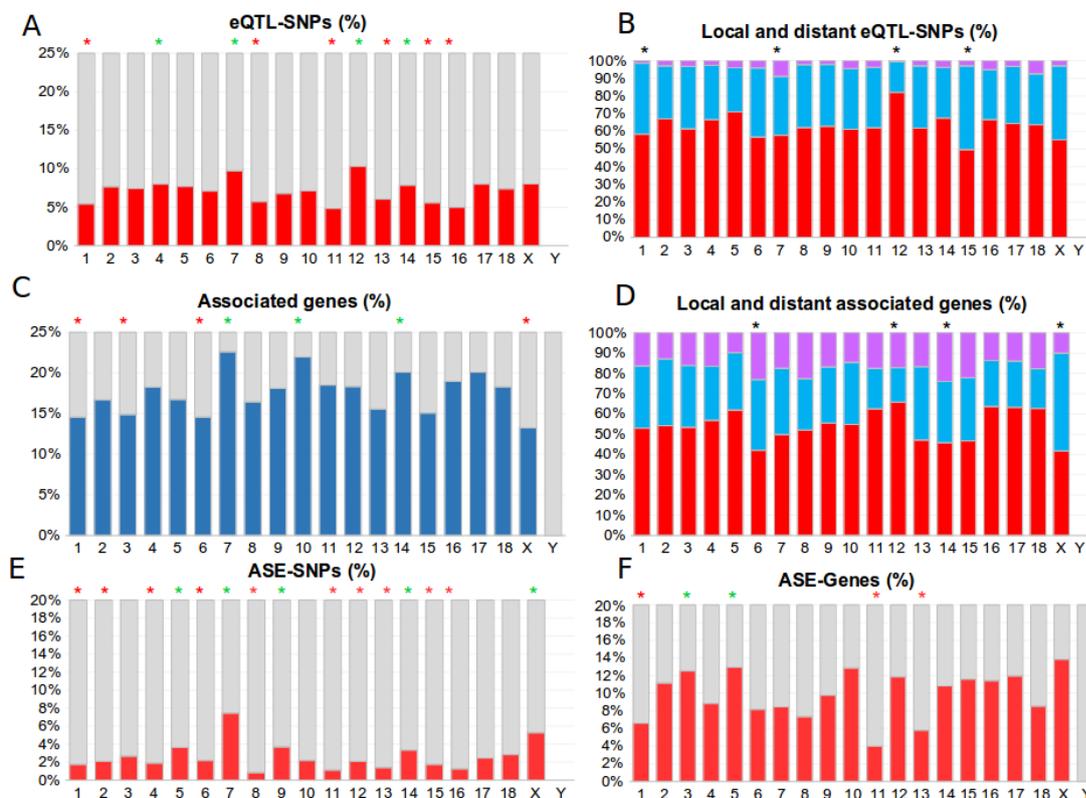


Figure 2



Figure 3: Biological processes enriched in genes associated with eQTL-SNPs (A) and/or affected by ASE (B)

A: The top 20 biological processes enriched in genes associated with eQTL-SNPs by eGWAS.

Horizontal bars represent the number of associated genes per biological process, with the corresponding q-values. Bars were split in order to represent the respective numbers of local-associated genes (dark red), distant-associated genes (blue), and local+distant-associated genes (purple).

B: Biological processes enriched in genes affected by ASE. The horizontal bars represent the number of ASE-genes involved in each biological process, with the corresponding q-values. The two enrichment analyses were performed with the GOrilla tool.

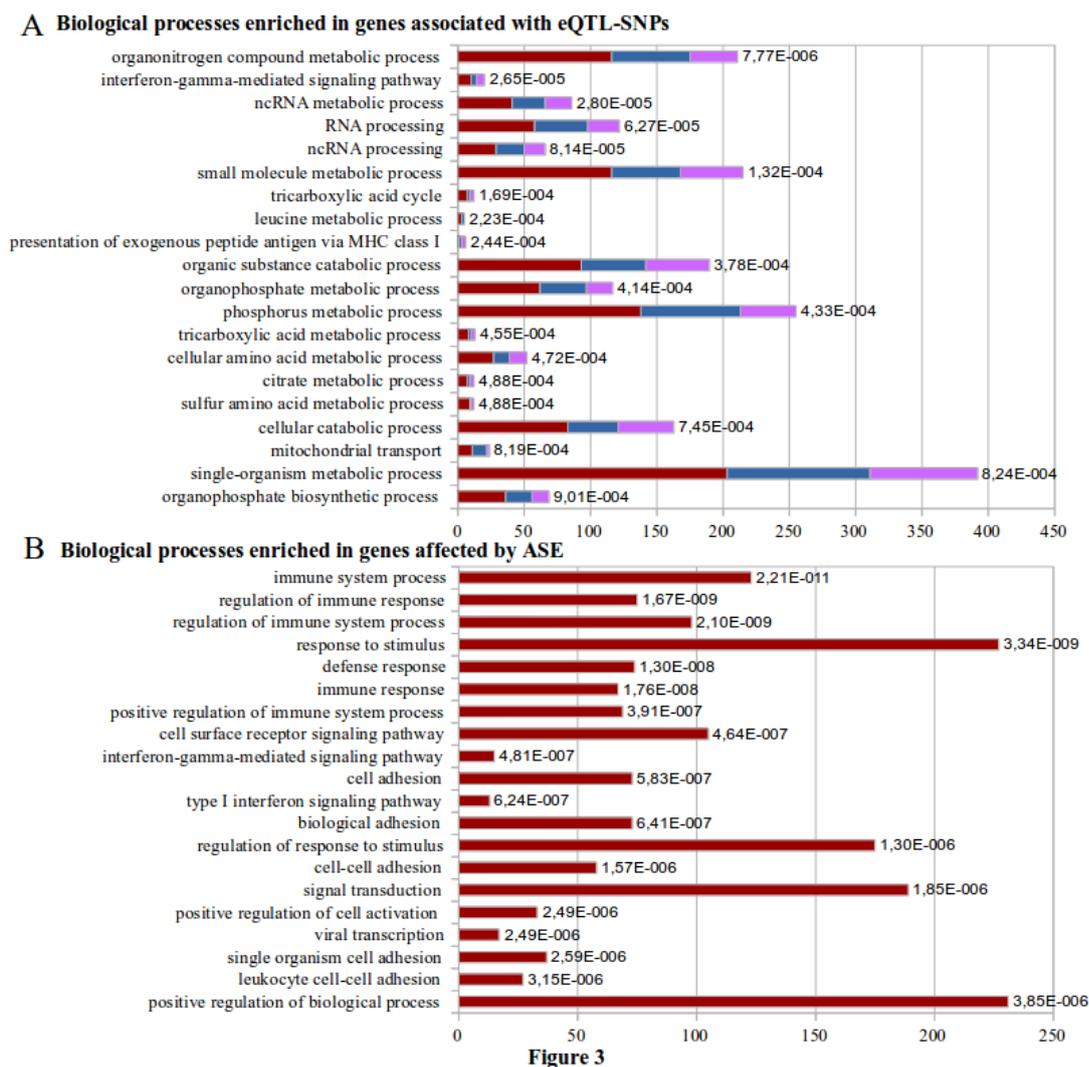


Figure 3

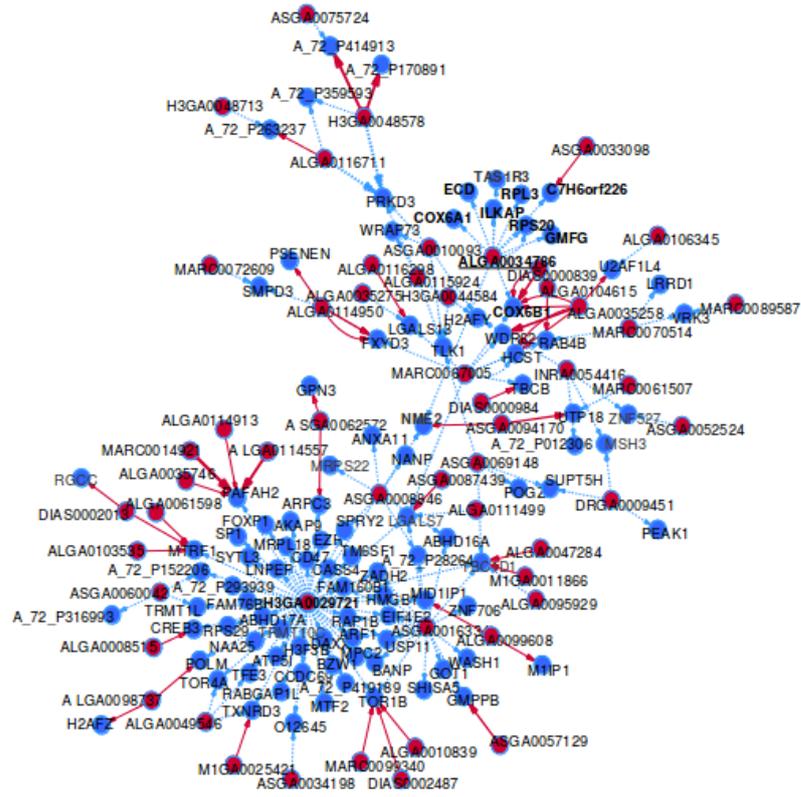


Figure 4: Visualization of association networks revealed by eGWAS (A, B) and links with ASE (C).

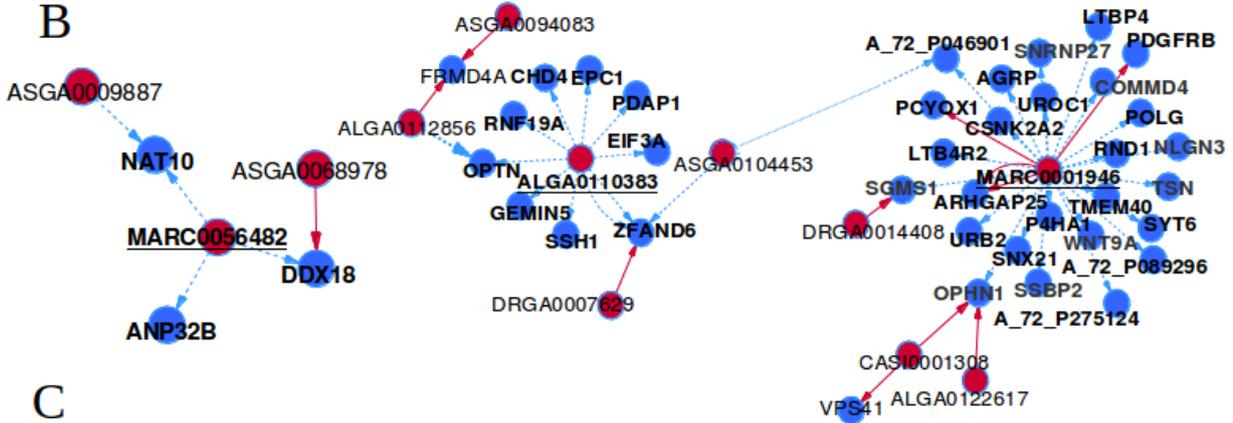
Association networks were constructed based on multiple associations between eQTL-SNPs or genes, and were visualized by Cytoscapev3.2.1. eQTL-SNPs and genes are represented by red and blue dots, respectively. Associations are depicted with directed edges (oriented arrows) to show regulation by the eQTL-SNP genotype on gene transcription and not the reverse. Local associations are drawn with red arrows and distant associations with blue dotted arrows. All nodes correspond either to a gene or an eQTL-SNP, with the names indicated. Names in bold are further described in the results. A: The biggest association network drawn in this study comprises 154 nodes (eQTL-SNPs or genes). It is centered on two eQTL-SNPs: H3GA0029721 (rs81422644), which was distant-associated with 51 probes representing 47 genes, and ALGA0034786 (rs81393122). B: Visualization of the association networks centered on eQTL-SNPs MARC0001946, ALGA0110383, and MARC0056482. C: Map of a segment of SSC6 that comprises the ASE-SNP rs242721226 and the eQTL-SNPs ALGA0110383 and MARC0056482. The distances between the position of ASE-SNP rs342721226 and the eQTL-SNPs ALGA0110383 and MARC0056482 are 676 and 381 Kb, respectively.



A



B



C

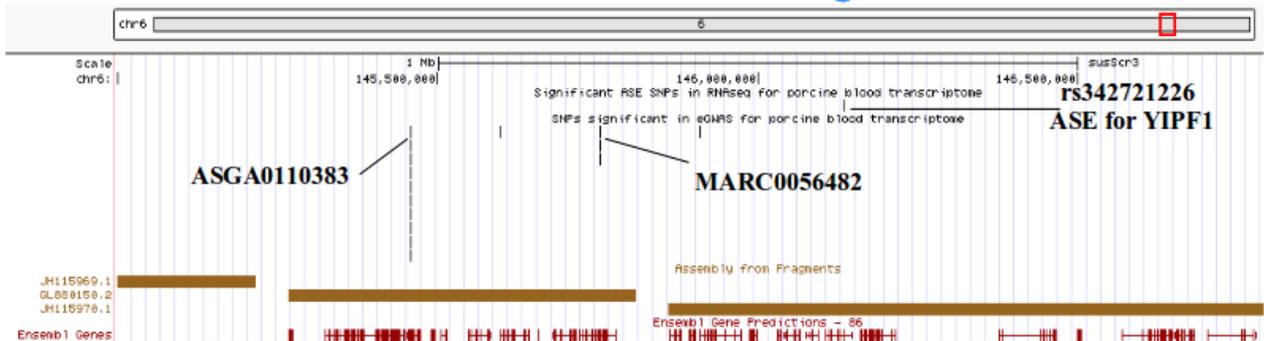


Figure 4



Figure 5: Detailed analysis of the association network centered on eQTL-SNP ALGA0034786 on SSC6

A: Circo plot produced by R package RCirco which maps eQTL-SNPs and associated genes according to their genomic positions. Links in red correspond to associations between the eQTL-SNP ALGA0034786 and genes mapping to SSC6, and links in blue represent associations with genes on other chromosomes. B: The distribution of all (grey) and significant (red) correlation coefficients calculated from pairwise comparisons of expression variation in probes associated with ALGA0034786, performed by PCIT. Red areas of the histogram correspond to coefficient correlations involving significant edges, as determined by PCIT, and gray areas represent non-significant edges. C: Boxplots of expression variations in probes annotated for porcine genes, depending on the eQTL-SNP genotype. Each dot represents one animal. Pink dots correspond to pigs homozygous for the most-frequent allele (genotype A/A), green dots to heterozygous pigs (genotype A/G), and blue dots to animals homozygous for the less-frequent allele (genotype G/G). D: Ingenuity Pathway Analysis of the genes included in the association network. The functional network comprised nine of the ten genes that were distant-associated to eQTL-SNP ALGA0034786. The symbols of these nine genes are colored in gray. The IPA-drawn network revealed functions linked to “cellular assembly and organization”, “cell death and survival”, and “endocrine system disorder”. E: eGWAS results for probe A_72_P211467, which was annotated for the gene *COX6B1*, visualized with a Manhattan plot. The association peak on SSC6 corresponds to the eQTL represented by eQTL-SNP ALGA0034786. The red horizontal line corresponds to an FDR threshold < 0.01 and the horizontal blue line to an FDR threshold < 0.05 .



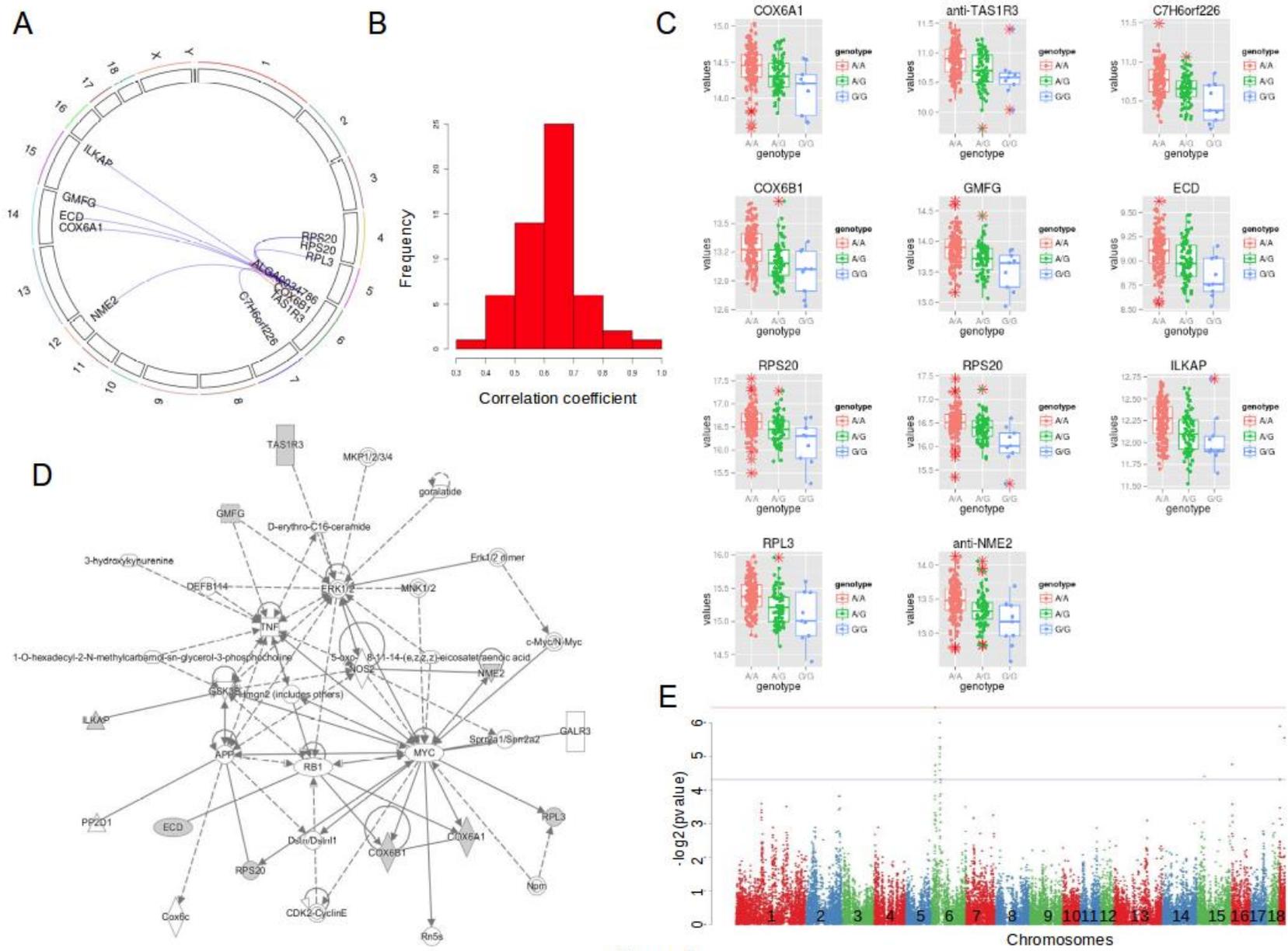


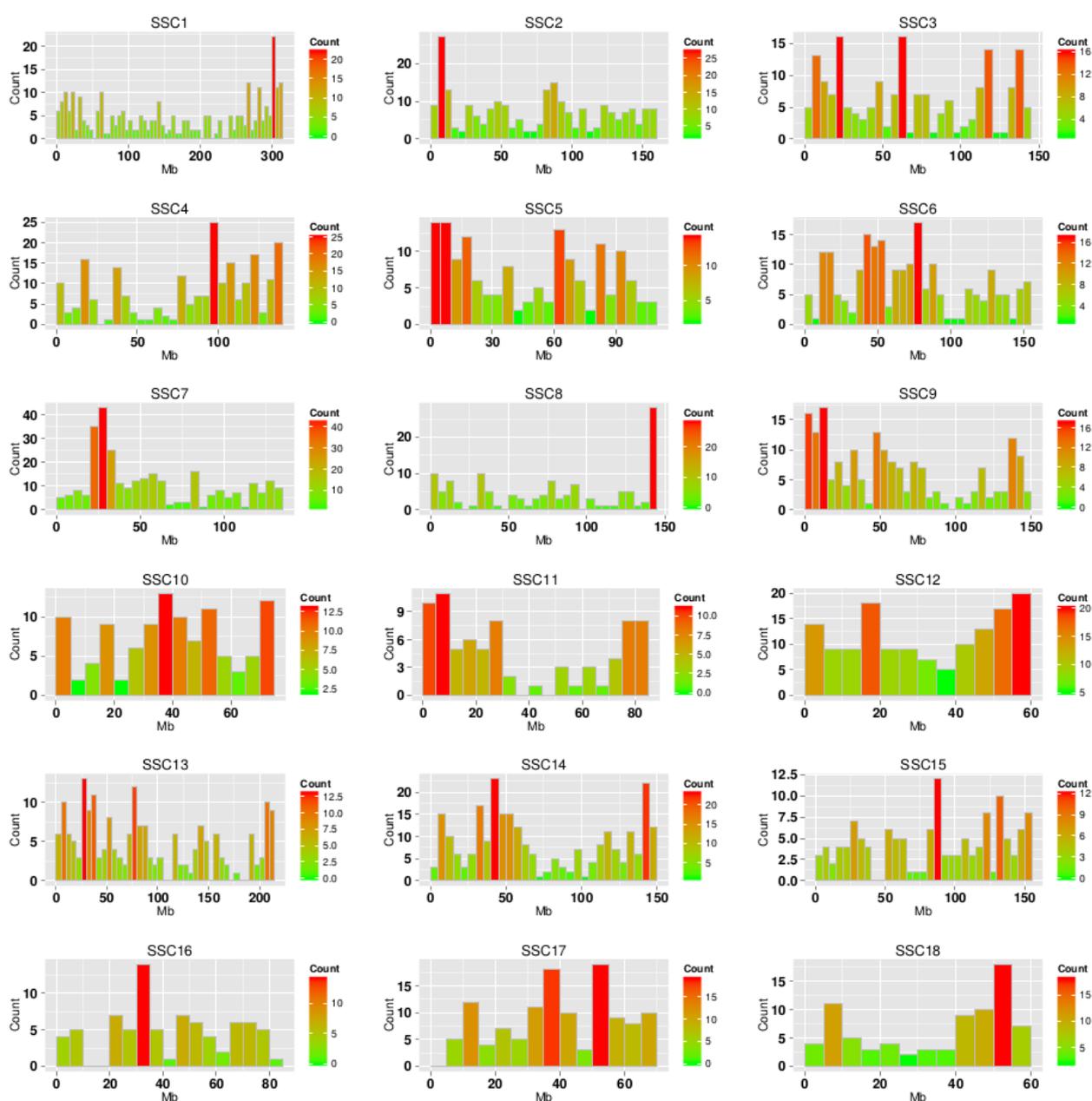
Figure 5



Supplementary files

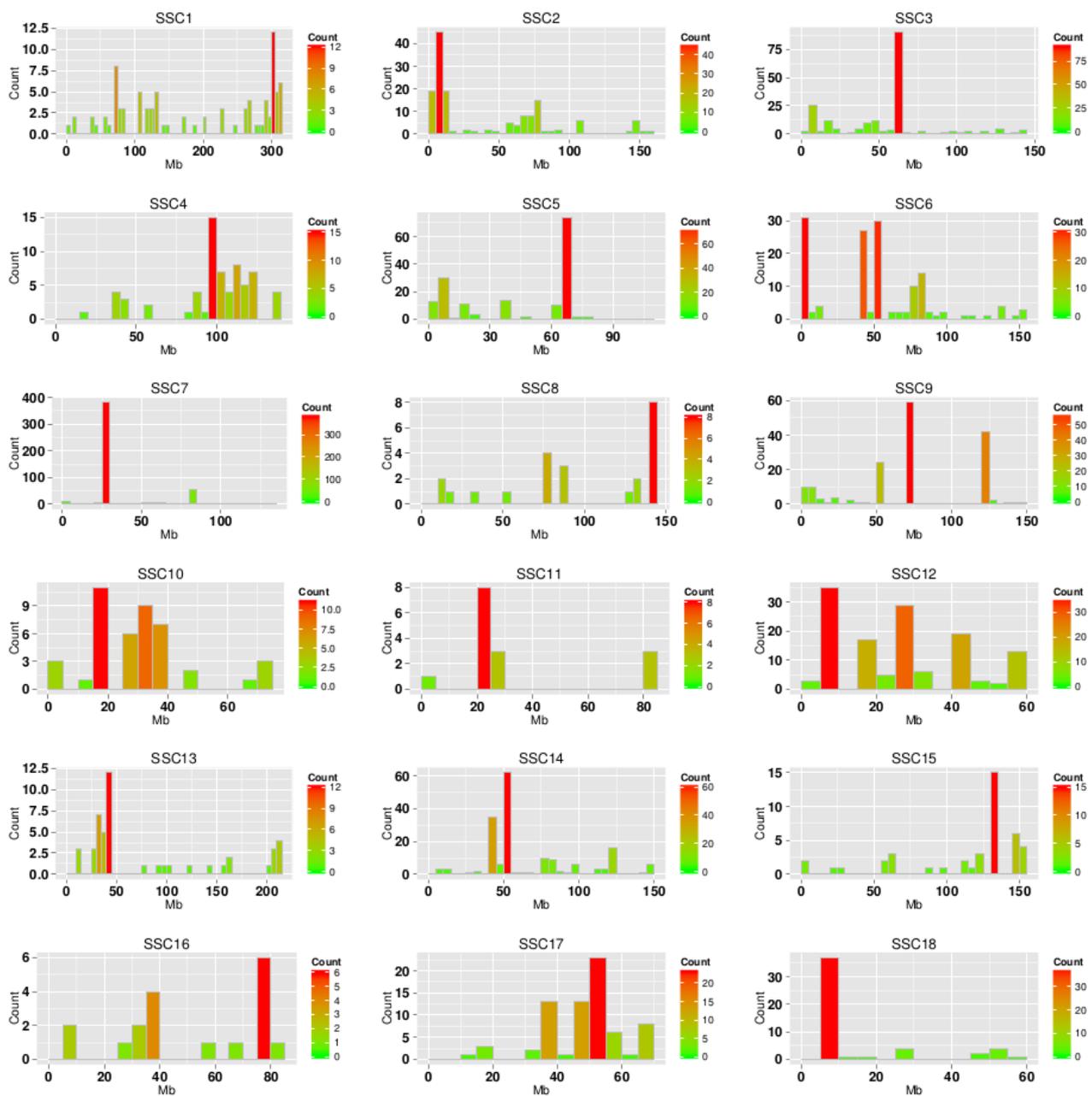
Supplementary Figure S1: Distribution of eQTL-SNPs along pig autosomal chromosomes

The histograms represent the distribution of eQTL-SNPs along chromosomes, with each bar corresponding to a 5-Mb segment. The density of eQTL-SNPs is represented by a color gradient from green to red. The scale for the eQTL-SNP count is on the Y-axis and varies among chromosomes. On SSC7, the region between 20 Mb and 30 Mb that has a high density of eQTL-SNPs overlaps the MHC locus.



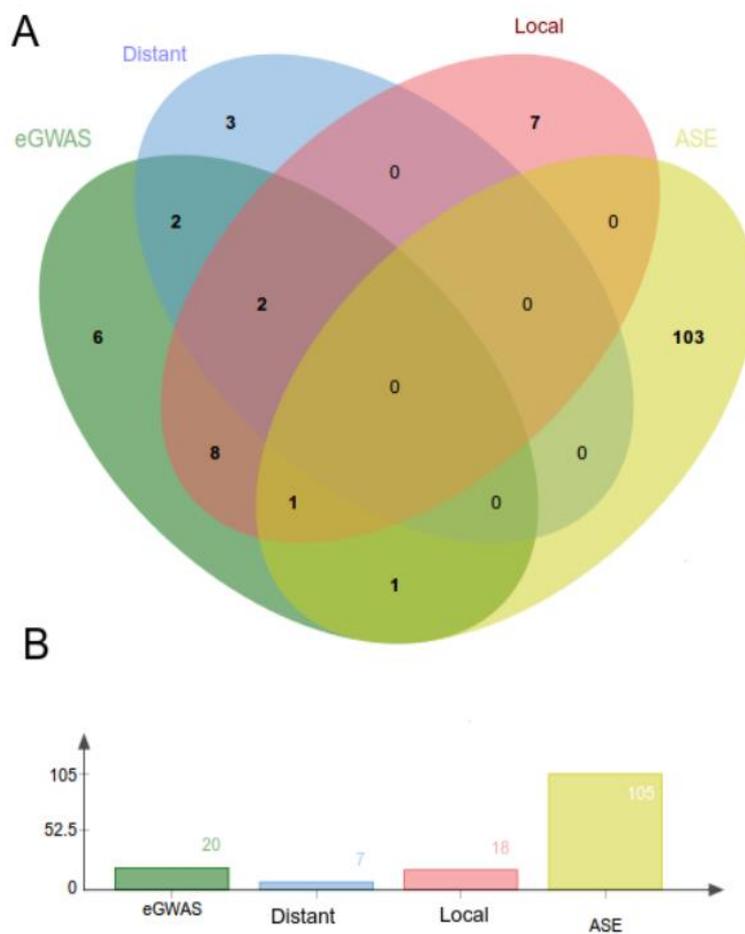
Supplementary Figure S2: Distribution of ASE-SNPs along pig autosomal chromosomes

The histograms represent the distribution of ASE-SNPs along chromosomes, with each bar corresponding to a 5-Mb segment. The density of ASE-SNPs is represented by a color gradient from green to red. The scale for the eQTL-SNP count is on the Y-axis and among chromosomes. The highest density of ASE-SNPs was found on SSC7 at the MHC locus.



Supplementary Figure S3: Comparison of enriched functions for genes detected by eGWAS (total, local, distant), and ASE analysis

A: Venn diagram produced with jvenn [54] showing shared and specific gene ontology (GO) terms. The numbers of enriched GO terms are shown for all associated genes detected by eGWAS (green), distant- (blue) or local-associated genes (red) detected by eGWAS, and ASE-genes (yellow). B: Number of enriched GO terms detected for each analysis.



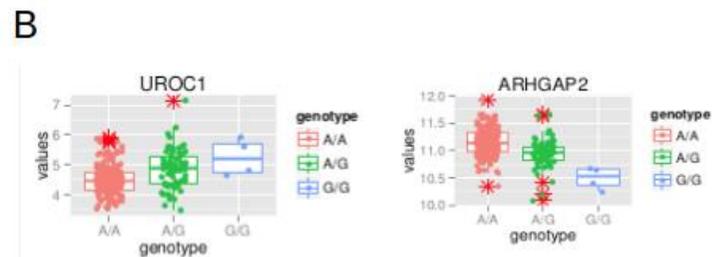
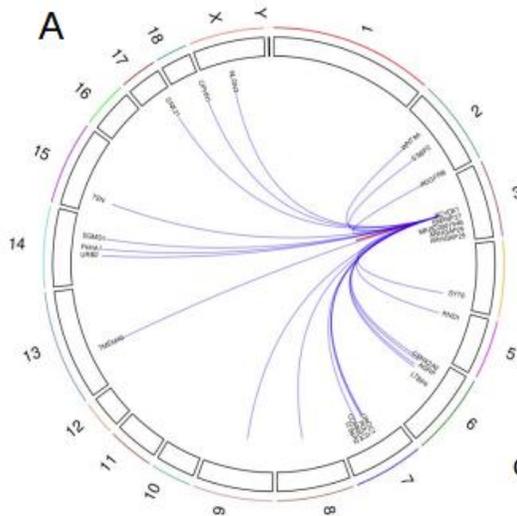
Supplementary Figure S3



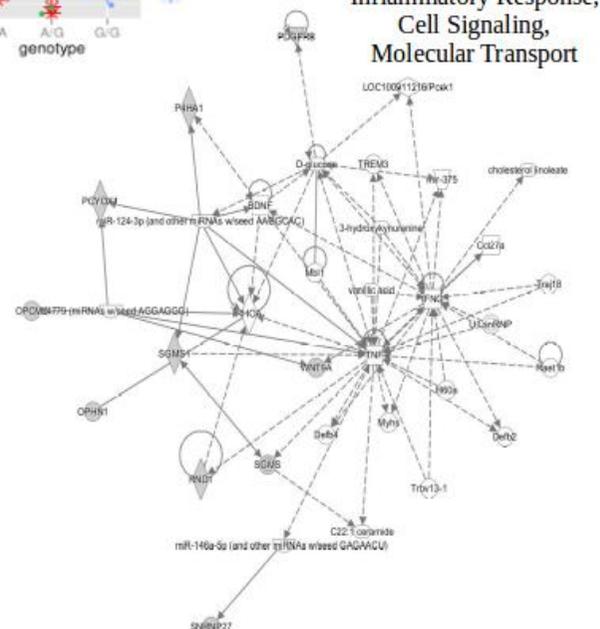
Supplementary Figure S5: Detailed analysis of the association network centered on eQTL-SNP MARC0001946

A: Circo plot produced by R package RCirco which maps eQTL-SNP MARC0001946 and its associated genes according to their genomic positions. Red links represent associations between MARC0001946 and genes on SSC3, and blue links represent associations with genes on other chromosomes. B: The boxplots represent the expression variation of probes annotated for the porcine genes *UROCI* and *ARHGAP2*, depending on eQTL-SNP genotype. Each dot represents one animal. Pink dots correspond to pigs homozygous for the more-frequent allele (genotype A/A), green dots to heterozygous pigs (genotype A/G), and blue dots to animals homozygous for the less-frequent allele (G/G). The boxplots show that the eQTL-SNP genotypes affected the transcription level of the two associated genes in opposite directions. C: The distribution of all (grey) and significant (red) correlation coefficients calculated from pairwise comparisons of expression variation in probes associated with MARC0001946, performed by the PCIT algorithm. D: Three functional networks produced by IPA, on which genes associated with eQTL-SNP MARC0001946 are represented by gray symbols. The functional network related to “cellular development”, “hematological system development and function”, and “hematopoiesis” includes ten associated genes; the network related to “inflammatory response”, “cell signaling”, and “molecular transport” includes nine associated genes; and the network related to “developmental disorder”, “hereditary disorder”, and “organismal injury and abnormalities” includes five associated genes.



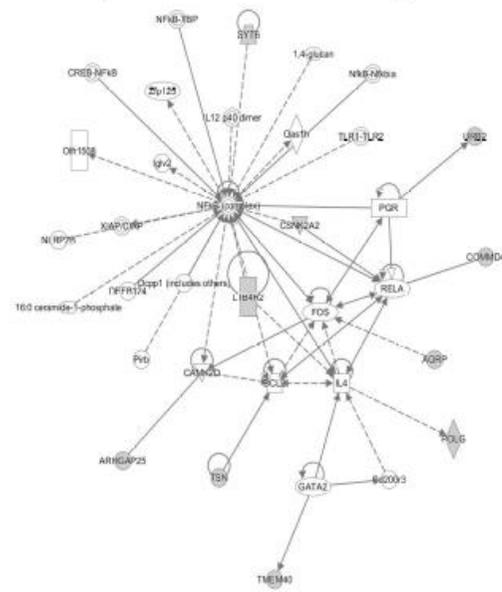


Inflammatory Response,
Cell Signaling,
Molecular Transport

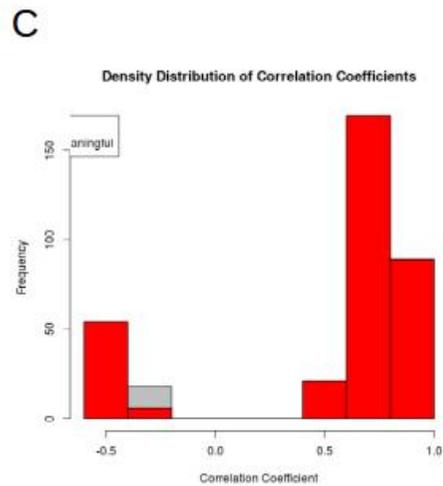
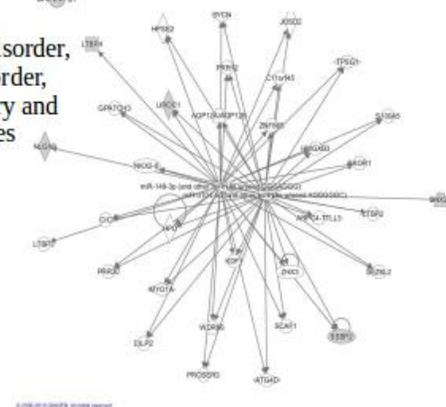


D

Cellular Development, Hematological System
Development and Function, Hematopoiesis



Developmental Disorder,
Hereditary Disorder,
Organismal Injury and
Abnormalities

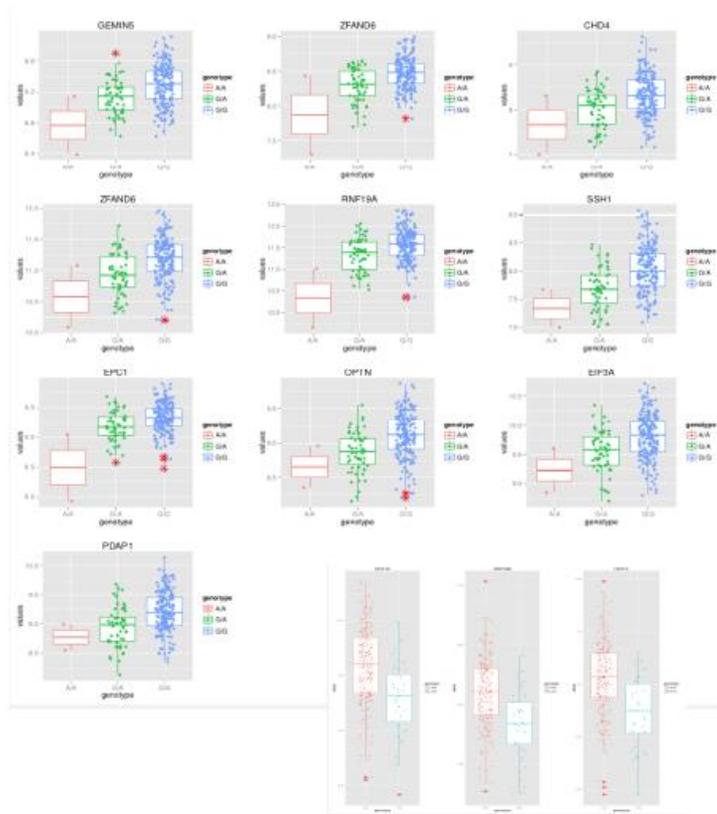
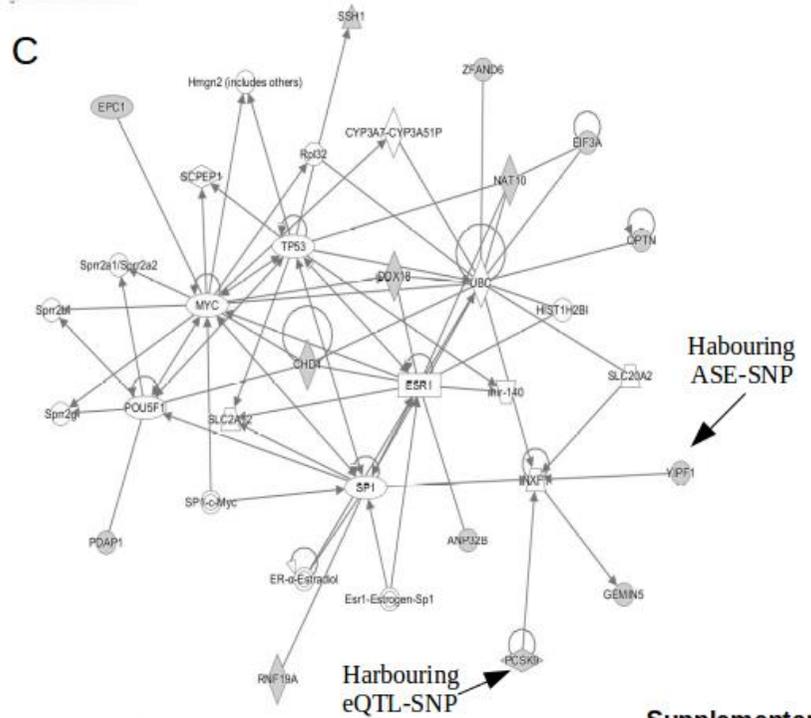
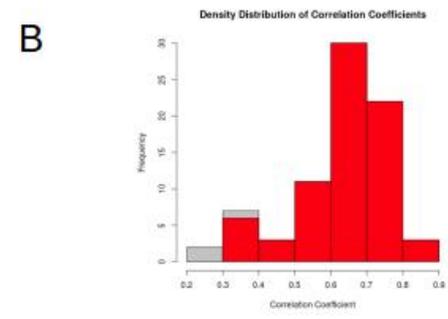
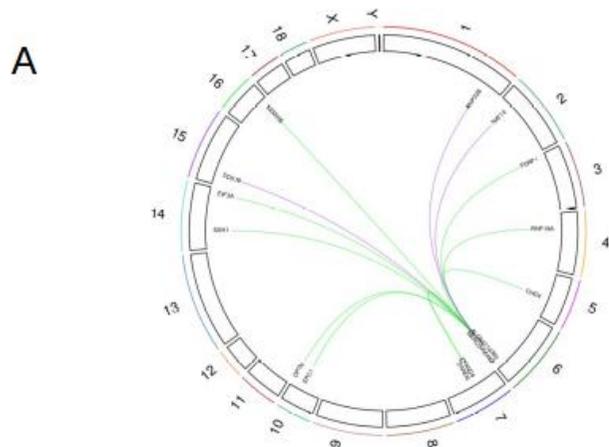


Supplementary Figure S5

Supplementary Figure S6: Detailed analysis of the association networks centered around eQTL-SNPs ALGA0110383 and MARC0056482

A: Circo plot produced by R package RCirco which maps the two eQTL-SNPs on SSC6 together with their associated genes, according to their genomic positions. Green links represent associations between genes and ALGA0110383. Purple links represent associations between genes and MARC0056482. B: The distribution of all (grey) and significant (red) correlation coefficients calculated from pairwise comparisons of expression variation in probes associated with ALGA0110383 and MARC0001946, performed by the PCIT algorithm. C: Functional network produced by IPA in which genes associated with eQTL-SNPs ALGA0110383 and MARC0001946, as well as the genes *YIPF1* and *PCSK9* (which harbor an ASE-SNP and an eQTL-SNP, respectively) are represented by gray-colored symbols. The functional network, which contains 14 associated genes, is related to “cell death and survival”, “organismal injury and abnormalities”, and “tissue morphology”. D: The boxplots represent expression variations in probes annotated for porcine genes according to each eQTL-SNP genotype. Each dot represents an animal. Pink dots correspond to pigs homozygous for the most-frequent allele (A/A), green dots to heterozygous pigs (A/G), and blue dots to animals homozygous for the less-frequent allele (G/G).





Supplementary Figure S6



Supplementary Table S1: Associations detected by eGWAS

SNP	SNP in dbSNP	SNP Position	Alleles	Associated Probe	Probe Position	Probe Strand	Assigned Gene	DE genes for Its*	Sens / Antisens Probe	Distance SNP-Probe (bp)	Association Type	Q-value	N	Allele Effect	MAF
H3GA0002262	rs80812787	1:100513906	C/A	A_72_P147961	1:99469729	+	HTR1B		Sens	1044177	distant	0,0035127	242	0,51	0,10
ALGA0000749	rs80831642	1:10070322	G/A	A_72_P709805	1:10050231	+	TAGAP		Sens	20091	local	0,0143868	239	-0,24	0,46
ASGA0003819	rs80800029	1:101141621	A/C	A_72_P367333	1:101074503	-	MYO6	3	Sens	67118	local	2,58E-11	242	1,03	0,38
MARC0068576	rs80891077	1:103959534	A/G	A_72_P330498	1:83701166	+	LACE1		Sens	20258368	distant	0,0017404	242	-0,5	0,07
ALGA0005006	rs80919228	1:104446745	A/G	A_72_P411398	13:37400825	-	DUSP7		Sens	NA	distant	0,0198086	242	0,43	0,07
ALGA0005086	rs81355341	1:106042541	G/A	A_72_P492340	1:106070192	+	C18orf25		Sens	27651	local	0,0399473	242	0,19	0,22
ASGA0000050	rs80907171	1:1067148	C/A	A_72_P489129	14:64382489	-	AGT		Sens	NA	distant	1,75E-19	239	-1,58	0,49
MARC0068314	rs80846545	1:108085908	A/G	A_72_P590729	9:40309507	-	SLN		Sens	NA	distant	0,0103364	242	0,44	0,08
ASGA0000934	rs81355262	1:10931163	G/A	O9455	1:10934602	+	SERAC1		Sens	3439	local	2,94E-07	242	0,29	0,48
ASGA0000934	rs81355262	1:10931163	G/A	A_72_P001396	1:10934621	+	SERAC1		Sens	3458	local	0,000025	242	0,31	0,48
CAIL0000072	rs327087023	1:110548193	G/A	A_72_P282494	1:110511509	+	SMAD4		Sens	36684	local	0,013257	242	-0,33	0,06
ALGA0114715	rs81343741	1:11181248	G/A	O9455	1:10934602	+	SERAC1		Sens	246646	local	0,0191886	242	-0,21	0,36
ASGA0097955	rs81318034	1:11234129	A/G	A_72_P257177	1:11517822	-	ZDHHC14		Sens	283693	local	0,0000449	242	0,28	0,43
ASGA0101578	rs81322709	1:11278294	A/G	A_72_P257177	1:11517822	-	ZDHHC14		Sens	239528	local	0,0112049	242	-0,24	0,32
CASI0009733	rs339166996	1:113658130	A/G	A_72_P212732	1:118842986	+			Non annotated	5184856	distant	0,0009228	242	1,15	0,21
ASGA0099868	-	1:11561288	A/G	A_72_P156986	15:23204426	-			Non annotated	NA	distant	0,0008202	242	0,18	0,24
ALGA0005334	rs81355500	1:115707215	A/G	A_72_P157951	1:116990000	-			Non annotated	1282785	distant	0,0017947	242	0,3	0,49
INRA0003647	rs321031460	1:117284820	A/G	A_72_P157951	1:116990000	-			Non annotated	294820	local	6,68E-10	242	0,57	0,19
INRA0003647	rs321031460	1:117284820	A/G	A_72_P189086	1:117627870	+			Non annotated	343050	local	1,03E-07	242	0,35	0,19
H3GA0002488	rs81355534	1:117846961	A/G	A_72_P507898	1:120133613	-	LACTB		Sens	2286652	distant	0,0301378	242	0,27	0,45
INRA0003669	rs322985995	1:118370336	A/G	A_72_P157951	1:116990000	-			Non annotated	1380336	distant	0,0012364	242	0,33	0,42
MARC0005036	rs80893317	1:119497071	G/A	A_72_P198352	1:119484798	+	HERC1		Sens	12273	local	1,2E-08	242	0,33	0,35



Supplementary Table S2. Genes annotated by sense and/or antisense probes found associated with local or distant eQTL-SNPs

	Local	Distant	Local+Distant	Total
Genes annotated with sense-probes	990	526	312	1828
Genes annotated with sense and antisense probes	29	5	39	73
Total	1019	531	351	1901
Genes annotated with antisense-probes	115	108	21	244
Non-annotated probes	220	147	54	421



Supplementary Table S3 : ASE analysis results

ASE-SNP	Chr	Position	Allele	Allele Freq 1	Allele Freq 2	Qual	Consequences	Impact	Genes	Symbols	DE Genes for Its*	Comparison with eGWAS**
1_10051601_T/G	1	10051601	T/G	T:0.960526	G:0.0394737	35.06	3_prime_UTR_variant	MODIFIER	ENSSSCG00000004053	TAGAP		1
1_10555288_T/C	1	10555288	T/C	T:0.954545	C:0.0454545	26.68	3_prime_UTR_variant	MODIFIER	ENSSSCG00000004494	-		
1_10555301_T/C	1	10555301	T/C	T:0.954545	C:0.0454545	24.27	3_prime_UTR_variant	MODIFIER	ENSSSCG00000004494	-		
rs80857024	1	108450589	T/A	T:0.684211	A:0.315789	15241.9	upstream_gene_variant; synonymous_variant	MODIFIER;LOW	ENSSSCG00000024372; ENSSSCG00000028329; ENSSSCG00000030194; ENSSSCG00000029642	SNORD58;RPL17		
1_108451392_C/T	1	108451392	C/T	C:0.945946	T:0.0540541	74.61	upstream_gene_variant	MODIFIER	ENSSSCG00000024372; ENSSSCG00000028329; ENSSSCG00000030194; ENSSSCG00000029642	SNORD58;RPL17		
1_108451393_T/C	1	108451393	T/C	T:0.945946	C:0.0540541	77.91	upstream_gene_variant	MODIFIER	ENSSSCG00000024372; ENSSSCG00000028329; ENSSSCG00000030194; ENSSSCG00000029642	SNORD58;RPL17		
rs336215448	1	117466539	C/T	C:0.815789	T:0.184211	43080.7	downstream_gene_variant	MODIFIER	ENSSSCG00000004541	FECH		
rs342908665	1	117483650	G/A	G:0.0882353	A:0.911765	9534.58	intron_variant	MODIFIER	ENSSSCG00000004541	FECH		
1_117509476_C/A	1	117509476	C/A	C:0.815789	A:0.184211	467.83	stop_gained	HIGH	ENSSSCG00000004542	NARS		
1_119998712_C/T	1	119998712	C/T	C:0.934211	T:0.0657895	86.61	3_prime_UTR_variant	MODIFIER	ENSSSCG00000026571	RAB8B		1
1_121965954_G/A	1	121965954	G/A	G:0.112903	A:0.887097	6955.31	non_coding_transcript_exon_variant,non_coding_transcript_variant	MODIFIER	ENSSSCG00000004575	-		
1_121966043_G/A	1	121966043	G/A	G:0.0416667	A:0.958333	27857.1	non_coding_transcript_exon_variant,non_coding_transcript_variant	MODIFIER	ENSSSCG00000004575	-		
1_125424330_A/G	1	125424330	A/G	A:0.833333	G:0.166667	259.43	missense_variant	MODERATE	ENSSSCG00000024660	-		
rs701070352	1	125424337	T/C	T:0.161765	C:0.838235	13999.3	missense_variant	MODERATE	ENSSSCG00000024660	-		
1_125424341_C/G	1	125424341	C/G	C:0.3	G:0.7	12620.1	synonymous_variant	LOW	ENSSSCG00000024660	-		
rs690189696	1	132221548	A/C	A:0.947368	C:0.0526316	22.17	3_prime_UTR_variant	MODIFIER	ENSSSCG00000004618	ARPP19		



Supplementary Table S4: Enrichment analysis of associated genes in eGWAS by Gorilla

GO Term	Description	P-value	FDR q-value	Enrichment t	N	B	n	b	Cis-genes	Trans-genes	Cis-trans genes
GO:1901564	organonitrogen compound metabolic process	7.77E-6	1.02E-1	1.30	8877	870	1653	211	36	20	13
GO:0060333	interferon-gamma-mediated signaling pathway	2.65E-5	1.73E-1	2.50	8877	43	1653	20	203	108	81
GO:0034660	ncRNA metabolic process	2.8E-5	1.22E-1	1.50	8877	307	1653	86	11	11	2
GO:0006396	RNA processing	6.27E-5	2.05E-1	1.38	8877	476	1653	122	83	38	42
GO:0034470	ncRNA processing	8.14E-5	2.13E-1	1.55	8877	228	1653	66	9	1	2
GO:0044281	small molecule metabolic process	1.32E-4	2.87E-1	1.25	8877	927	1653	215	7	2	3
GO:0006099	tricarboxylic acid cycle	1.69E-4	3.15E-1	2.93	8877	22	1653	12	27	12	13
GO:0006551	leucine metabolic process	2.23E-4	3.64E-1	5.37	8877	5	1653	5	8	2	3
GO:0002480	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent	2.44E-4	3.54E-1	4.60	8877	7	1653	6	138	75	42
GO:1901575	organic substance catabolic process	3.78E-4	4.94E-1	1.24	8877	821	1653	190	62	35	20
GO:0019637	organophosphate metabolic process	4.14E-4	4.92E-1	1.33	8877	473	1653	117	93	49	48
GO:0006793	phosphorus metabolic process	4.33E-4	4.71E-1	1.20	8877	1143	1653	255	1	2	3
GO:0072350	tricarboxylic acid metabolic process	4.55E-4	4.58E-1	2.59	8877	27	1653	13	3	2	0
GO:0006520	cellular amino acid metabolic process	4.72E-4	4.4E-1	1.55	8877	180	1653	52	7	2	3
GO:0006101	citrate metabolic process	4.88E-4	4.25E-1	2.69	8877	24	1653	12	116	52	47
GO:0000096	sulfur amino acid metabolic process	4.88E-4	3.99E-1	2.69	8877	24	1653	12	29	21	16
GO:0044248	cellular catabolic process	7.45E-4	5.73E-1	1.25	8877	700	1653	163	58	40	24
GO:0006839	mitochondrial transport	8.19E-4	5.94E-1	1.90	8877	68	1653	24	41	25	20
GO:0044710	single-organism metabolic process	8.24E-4	5.67E-1	1.14	8877	1848	1653	392	10	4	6
GO:0090407	organophosphate biosynthetic process	9.01E-4	5.88E-1	1.43	8877	260	1653	69	116	59	36



Supplementary Table S5: Enrichment analysis of local-associated genes in eGWAS by Gorilla

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0044281	small molecule metabolic process	5.85E-5	7.64E-1	1.32	8878	927	1183	163
GO:0060333	interferon-gamma-mediated signaling pathway	6.89E-5	4.5E-1	2.79	8878	43	1183	16
GO:0006790	sulfur compound metabolic process	7.62E-5	3.32E-1	1.75	8878	197	1183	46
GO:0000096	sulfur amino acid metabolic process	1.06E-4	3.47E-1	3.44	8878	24	1183	11
GO:1901564	organonitrogen compound metabolic process	1.48E-4	3.86E-1	1.31	8878	870	1183	152
GO:0044248	cellular catabolic process	2.39E-4	5.2E-1	1.34	8878	700	1183	125
GO:0006099	tricarboxylic acid cycle	2.4E-4	4.47E-1	3.41	8878	22	1183	10
GO:0006082	organic acid metabolic process	2.41E-4	3.94E-1	1.40	8878	519	1183	97
GO:0072350	tricarboxylic acid metabolic process	3.79E-4	5.5E-1	3.06	8878	27	1183	11
GO:0043436	oxoacid metabolic process	5.27E-4	6.89E-1	1.38	8878	511	1183	94
GO:0006101	citrate metabolic process	5.65E-4	6.71E-1	3.13	8878	24	1183	10
GO:1901575	organic substance catabolic process	5.66E-4	6.16E-1	1.29	8878	821	1183	141
GO:0006520	cellular amino acid metabolic process	6.42E-4	6.46E-1	1.67	8878	180	1183	40
GO:2000786	positive regulation of autophagosome assembly	6.93E-4	6.47E-1	5.36	8878	7	1183	5
GO:0044090	positive regulation of vacuole organization	7.19E-4	6.26E-1	4.50	8878	10	1183	6
GO:0034660	ncRNA metabolic process	7.23E-4	5.9E-1	1.49	8878	307	1183	61
GO:0030490	maturation of SSU-rRNA	8.3E-4	6.38E-1	3.00	8878	25	1183	10
GO:0070126	mitochondrial translational termination	8.35E-4	6.06E-1	2.31	8878	52	1183	16



Supplementary Figure S6: Enrichment analysis of distant-associated genes in eGWAS by Gorilla

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0034470	ncRNA processing	1.17E-4	1	1.88	8878	228	765	37
GO:0006396	RNA processing	1.9E-4	1	1.56	8878	476	765	64
GO:0042118	endothelial cell activation	2.55E-4	1	9.28	8878	5	765	4
GO:0034660	ncRNA metabolic process	2.62E-4	8.54E-1	1.70	8878	307	765	45
GO:0044282	small molecule catabolic process	4.09E-4	1	1.88	8878	191	765	31
GO:1901575	organic substance catabolic process	6.13E-4	1	1.37	8878	821	765	97
GO:0090234	regulation of kinetochore assembly	6.38E-4	1	11.61	8878	3	765	3



Supplementary Table S7: Enrichment analysis of cis-genes in ASE analysis by Gorilla

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0002376	immune system process	2.21E-11	2.46E-7	1.76	5903	712	578	123
GO:0050776	regulation of immune response	1.67E-9	9.28E-6	1.99	5903	385	578	75
GO:0002682	regulation of immune system process	2.1E-9	7.78E-6	1.78	5903	561	578	98
GO:0050896	response to stimulus	3.34E-9	9.31E-6	1.38	5903	1686	578	227
GO:0006952	defense response	1.3E-8	2.89E-5	1.91	5903	395	578	74
GO:0006955	immune response	1.76E-8	3.26E-5	1.98	5903	346	578	67
GO:0002684	positive regulation of immune system process	3.91E-7	6.22E-4	1.81	5903	389	578	69
GO:0007166	cell surface receptor signaling pathway	4.64E-7	6.46E-4	1.58	5903	680	578	105
GO:0060333	interferon-gamma-mediated signaling pathway	4.81E-7	5.96E-4	4.26	5903	36	578	15
GO:0007155	cell adhesion	5.83E-7	6.5E-4	1.76	5903	424	578	73
GO:0060337	type I interferon signaling pathway	6.24E-7	6.32E-4	4.74	5903	28	578	13
GO:0022610	biological adhesion	6.41E-7	5.95E-4	1.75	5903	425	578	73
GO:0048583	regulation of response to stimulus	1.3E-6	1.11E-3	1.36	5903	1312	578	175
GO:0098609	cell-cell adhesion	1.57E-6	1.25E-3	1.86	5903	319	578	58
GO:0007165	signal transduction	1.85E-6	1.37E-3	1.33	5903	1448	578	189
GO:0050867	positive regulation of cell activation	2.49E-6	1.73E-3	2.32	5903	145	578	33
GO:0019083	viral transcription	2.49E-6	1.63E-3	3.47	5903	50	578	17
GO:0098602	single organism cell adhesion	2.59E-6	1.61E-3	2.20	5903	172	578	37
GO:0007159	leukocyte cell-cell adhesion	3.15E-6	1.85E-3	2.55	5903	108	578	27
GO:0048518	positive regulation of biological process	3.85E-6	2.15E-3	1.27	5903	1861	578	231



Supplementary Table S8: QTLs in AnimalQTLdb (<10Mb) for pigs harbouring at least one eQTL-SNP

Chromosome	Start position QTL	End position QTL	Phenotypes
1	3008501	3977158	Intramuscular fat content QTL (29590)
1	6986588	6986628	Backfat at rump QTL (23304)
1	7939563	8069663	Drip loss QTL (7867)
1	7939563	8069663	Drip loss QTL (7899)
1	7939563	8069663	Drip loss QTL (8030)
1	7939563	8069663	Drip loss QTL (8147)
1	7954082	8084181	Average backfat thickness QTL (267)
1	7964370	10707403	Musculus gluteus weight QTL (1276)
1	7964370	16114132	Melting point QTL (3123)
1	7964370	16114132	Palmitoleic acid content QTL (3109)
1	7964370	16114132	Palmitoleic acid content QTL (3118)
1	7964370	16114132	Stearic acid content QTL (3111)
1	7964370	16114132	Stearic acid content QTL (3119)
1	8715201	12868209	CIE-a* QTL (28043)
1	8715201	12868209	CIE-b* QTL (28044)
1	8715201	12868209	CIE-L* QTL (28045)
1	9246659	9336595	Drip loss QTL (8053)
1	9246659	9336595	Drip loss QTL (8338)
1	9246659	9336595	Drip loss QTL (8598)
1	9324566	9324606	Residual feed intake QTL (31336)
1	9999863	10264707	Drip loss QTL (7871)
1	9999863	10264707	Drip loss QTL (7896)
1	9999863	10264707	Drip loss QTL (8587)
1	10707403	16114132	average daily gain QTL (319)



Supplementary Table S9: QTLs in AnimalQTLdb for pigs (<10Mb) harbouring at least one ASE-SNPs

Chromosome	Start position QTL	End position QTL	Phenotype
1	3008501	3977158	Intramuscular fat content QTL (29590)
1	7964370	10707403	Musculus gluteus weight QTL (1276)
1	7964370	16114132	Melting point QTL (3123)
1	7964370	16114132	Palmitoleic acid content QTL (3109)
1	7964370	16114132	Palmitoleic acid content QTL (3118)
1	7964370	16114132	Stearic acid content QTL (3111)
1	7964370	16114132	Stearic acid content QTL (3119)
1	8715201	12868209	CIE-a* QTL (28043)
1	8715201	12868209	CIE-b* QTL (28044)
1	8715201	12868209	CIE-L* QTL (28045)
1	9999863	10264707	Drip loss QTL (7871)
1	9999863	10264707	Drip loss QTL (7896)
1	9999863	10264707	Drip loss QTL (8587)
1	10707403	16114132	average daily gain QTL (319)
1	10707403	16114132	Salmonella count in liver QTL (11554)
1	14013581	14153111	Drip loss QTL (7827)
1	14013581	14153111	Drip loss QTL (7829)
1	14013581	14153111	Drip loss QTL (7994)
1	14013581	14153111	Drip loss QTL (8024)
1	14013581	14153111	Drip loss QTL (8034)
1	35026068	35987086	Average backfat thickness QTL (29564)
1	35046144	35987086	CIE-a* QTL (95258)



Supplementary Table S10: Complete Blood Cell Counts for 195 animals among the 243 eGWAS pigs

	White cells (cells/mm ³)	Lymphocytes (% of white cells)	Monocytes (% of white cells)	Granulocytes (% of white cells)	EOS (% of white cells)	Red cells (cells/mm ³)	Hemoglobin (g/dl)	Hematocrit (% red cells of blood cells)	Mean cell volum (fl)
Mean	19,28	31,07	7,92	61,01	5,19	6,7	10,76	35,36	52,9
Median	14	30,8	7,9	61,1	4,9	6,64	10,59	34,81	53
Standard deviation	3,78	4,67	1,13	5,32	1,79	0,86	1,63	4,69	3,4
Variation coefficient	0,20	0,15	0,14	0,09	0,34	0,13	0,15	0,13	0,06



References

1. Clapperton M, Diack AB, Matika O, Glass EJ, Gladney CD, Mellencamp MA, *et al.* Traits associated with innate and adaptive immunity in pigs: heritability and associations with performance under different health status conditions. *Genetics Selection Evolution*. 2009;41:54.
2. Flori L, Gao Y, Laloë D, Lemonnier G, Leplat J-J, Teillaud A, *et al.* Immunity traits in pigs: substantial genetic variation and limited covariation. *PLoS ONE*. 2011;6:e22717.
3. Clapperton M, Glass EJ, Bishop SC. Pig peripheral blood mononuclear leucocyte subsets are heritable and genetically correlated with performance. *Animal*. 2008;2:1575–84.
4. Mach N, Gao Y, Lemonnier G, Lecardonnel J, Oswald IP, Estellé J, *et al.* The peripheral blood transcriptome reflects variations in immunity traits in swine: towards the identification of biomarkers. *BMC Genomics*. 2013;14:894.
5. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet*. 2001;29:389–95.
6. Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat. Rev. Genet*. 2006;7:862–72.
7. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature*. 2003;422:297–302.
8. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet*. 2015;16:197–212.
9. Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet*. 2010;11:533–8.
10. Lagarrigue S, Martin L, Hormozdiari F, Roux P-F, Pan C, van Nas A, *et al.* Analysis of allele-specific expression in mouse liver by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. *Genetics*. 2013;195:1157–66.
11. Hasin-Brumshtein Y, Hormozdiari F, Martin L, van Nas A, Eskin E, Lusis AJ, *et al.* Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics*. 2014;15:471.
12. Joehanes R, Zhang X, Huan T, Yao C, Ying S-X, Nguyen QT, *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol*. 2017;18:16.
13. Liaubet L, Lobjois V, Faraut T, Tircazes A, Benne F, Iannuccelli N, *et al.* Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. *BMC Genomics*. 2011;12:548.
14. Martínez-Montes AM, Muiños-Bühl A, Fernández A, Folch JM, Ibáñez-Escriche N, Fernández AI. Deciphering the regulation of porcine genes influencing growth, fatness and yield-related traits through genetical genomics. *Mamm. Genome*. 2016;



15. Ponsuksili S, Murani E, Trakooljul N, Schwerin M, Wimmers K. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. *Int. J. Biol. Sci.* 2014;10:327–37.
16. Cinar MU, Kayan A, Uddin MJ, Jonas E, Tesfaye D, Phatsara C, *et al.* Association and expression quantitative trait loci (eQTL) analysis of porcine AMBP, GC and PPP1R3B genes with meat quality traits. *Mol. Biol. Rep.* 2012;39:4809–21.
17. González-Prendes R, Quintanilla R, Cánovas A, Manunza A, Figueiredo Cardoso T, Jordana J, *et al.* Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Sci Rep.* 2017;7:39830.
18. Ponsuksili S, Murani E, Schwerin M, Schellander K, Wimmers K. Identification of expression QTL (eQTL) of genes expressed in porcine *M. longissimus dorsi* and associated with meat quality traits. *BMC Genomics.* 2010;11:572.
19. Puig-Oliveras A, Revilla M, Castelló A, Fernández AI, Folch JM, Ballester M. Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci Rep.* 2016;6:31803.
20. Ponsuksili S, Zebunke M, Murani E, Trakooljul N, Krieter J, Puppe B, *et al.* Integrated Genome-wide association and hypothalamus eQTL studies indicate a link between the circadian rhythm-related gene *PER1* and coping behavior. *Sci Rep.* 2015;5:16264.
21. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:122.
22. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, *et al.* Ensembl 2016. *Nucleic Acids Res.* 2016;44:D710-716.
23. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009;10:48.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
25. Watson-Haigh NS, Kadarmideen HN, Reverter A. PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics.* 2010;26:411–3.
26. Reverter A, Chan EKF. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics.* 2008;24:2491–7.
27. Fehrmann RSN, Jansen RC, Veldink JH, Westra H-J, Arends D, Bonder MJ, *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* 2011;7:e1002197.
28. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* 2012;44:502–10.



29. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science*. 2014;343:1246949.
30. Renard C, Hart E, Sehra H, Beasley H, Coggill P, Howe K, *et al.* The genomic sequence and analysis of the swine major histocompatibility complex. *Genomics*. 2006;88:96–110.
31. Hu Z-L, Fritz ER, Reecy JM. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucl. Acids Res*. 2007;35:D604–9.
32. Schramm K, Marzi C, Schurmann C, Carstensen M, Reinmaa E, Biffar R, *et al.* Mapping the genetic architecture of gene regulation in whole blood. *PLoS ONE*. 2014;9:e93844.
33. Sasayama D, Hori H, Nakamura S, Miyata R, Teraishi T, Hattori K, *et al.* Identification of single nucleotide polymorphisms regulating peripheral blood mRNA expression with genome-wide significance: an eQTL study in the Japanese population. *PLoS ONE*. 2013;8:e54967.
34. Mehta D, Heim K, Herder C, Carstensen M, Eckstein G, Schurmann C, *et al.* Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur. J. Hum. Genet*. 2013;21:48–54.
35. Rotival M, Zeller T, Wild PS, Maouche S, Szymczak S, Schillert A, *et al.* Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet*. 2011;7:e1002367.
36. Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, *et al.* Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE*. 2009;4:e6524.
37. Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H, *et al.* Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet*. 2006;2:e172.
38. Landmark-Høyvik H, Dumeaux V, Nebdal D, Lund E, Tost J, Kamatani Y, *et al.* Genome-wide association study in breast cancer survivors reveals SNPs associated with gene expression of genes belonging to MHC class I and II. *Genomics*. 2013;102:278–87.
39. Wong D, Lee W, Humburg P, Makino S, Lau E, Naranbhai V, *et al.* Genomic mapping of the MHC transactivator CIITA using an integrated ChIP-seq and genetical genomics approach. *Genome Biol*. 2014;15:494.
40. Schroyen M, Tuggle CK. Current transcriptomics in pig immunity research. *Mamm. Genome*. 2015;26:1–20.
41. Chen C, Yang B, Zeng Z, Yang H, Liu C, Ren J, *et al.* Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. *BMC Genomics*. 2013;14:848.
42. Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G, *et al.* Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J*. 2016;10:2973–7.



43. Mach N, Berri M, Estellé J, Levenez F, Lemonnier G, Denis C, *et al.* Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ Microbiol Rep.* 2015;7:554–69.
44. Sanchez M-P, Tribout T, Iannuccelli N, Bouffaud M, Servin B, Tenghe A, *et al.* A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. *Genet. Sel. Evol.* 2014;46:12.
45. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* 2007;23:1294–6.
46. Jacquier V, Estellé J, Schmaltz-Panneau B, Lecardonnel J, Moroldo M, Lemonnier G, *et al.* Genome-wide immunity studies in the rabbit: transcriptome variations in peripheral blood mononuclear cells after in vitro stimulation by LPS or PMA-Ionomycin. *BMC Genomics.* 2015;16:26.
47. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
48. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25:1105–11.
49. Chen W-M, Abecasis GR. Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* 2007;81:913–26.
50. Bryois J, Buil A, Evans DM, Kemp JP, Montgomery SB, Conrad DF, *et al.* Cis and Trans Effects of Human Genomic Variants on Gene Expression. *PLOS Genetics.* 2014;10:e1004461.
51. Mach N, Berri M, Esquerré D, Chevaleyre C, Lemonnier G, Billon Y, *et al.* Extensive expression differences along porcine small intestine evidenced by transcriptome sequencing. *PLoS ONE.* 2014;9:e88515.
52. DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
53. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics.* 2013;14:244.
54. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics.* 2014;15:293.



II. Etude 2 : Etude des profils transcriptomiques de quatre tissus immunitaires

Article soumis à *Scientific Reports*

A. Objectifs et approches

Cette étude a porté sur les Plaques de Peyer et le ganglion mésentérique qui constituent les tissus lymphoïdes secondaires associés au système digestif (GALT pour *Gut Associated Lymphoid Tissues* en anglais) et le sang périphérique, tissu circulant où transitent de nombreux acteurs de l'immunité.

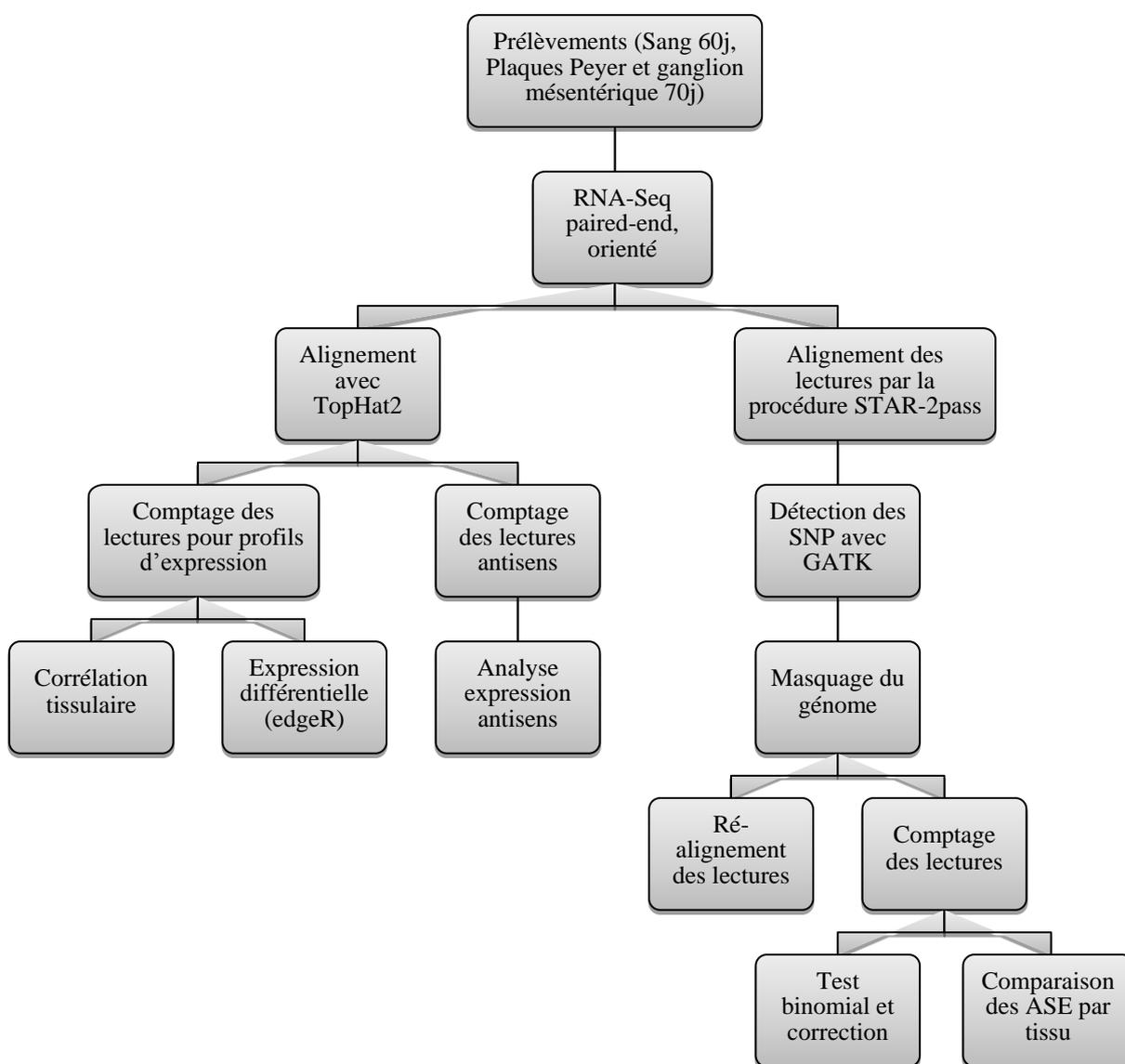


Figure E2-27 : Workflow d'analyses suivi pour l'étude 2



Les objectifs de cette étude étaient : (i) Etudier les profils transcriptomiques de chaque tissu, (ii) Evaluer la corrélation entre les profils d'expression, (iii) Identifier des gènes différentiellement exprimés entre les tissus, (iv) Etudier le contrôle génétique *cis* de l'expression, (vi) Mettre en évidence des ASE spécifiques de chaque tissu.

L'étude a été réalisée sur quatre mâles Large-White non castrés produits lors du projet SUS_FLORA. Le sang a été prélevé à 60 jours. Les plaques de Peyer et le ganglion mésentérique ont été prélevés à l'abattage des animaux à 70 jours. Tous les échantillons ont été analysés par RNA-Seq avec un protocole permettant de séquencer les ARN de manière orientée. Les analyses sont représentées sur la Figure E2-27.

B. Résultats

Dans cette étude, l'expression de plus de 13 000 gènes a été estimée dans le ganglion mésentérique et les deux types de plaques de Peyer et 9 000 gènes pour le sang. Plus de 3 000 gènes ont révélé des niveaux de transcription antisens mesurables pour le GALT et 1 000 pour le sang. Parmi ces gènes, environ 300 ne présentaient pas de transcription dans le sens pour lesquels ils sont annotés.

La corrélation d'expression des gènes entre les deux types de plaques de Peyer est égale à 0,97. Elle est supérieure à 0,8 entre les plaques de Peyer et le ganglion mésentérique et entre le ganglion mésentérique et le sang et elle est proche de 0,7 entre le sang et les plaques de Peyer.

Nous avons identifié 1 103 gènes différentiellement exprimés (DE) entre les Plaques de Peyer iléale et jéjunale et six fois plus de gènes DE entre les plaques de Peyer et le ganglion mésentérique. Les régulateurs de transcription FOXP3, GATA3, STAT4, TBX21 (qui code pour T-bet) et RORC (qui code pour Ror γ t) ont été trouvés sous-exprimés dans la plaque de Peyer iléale par rapport à la plaque de Peyer jéjunale, alors que le facteur BCL6 s'est avéré sur-exprimé dans la plaque de Peyer iléale.

Nous avons trouvé une transcription antisens pour 24%, 22% et 14% des transcrits détectés respectivement dans le ganglion mésentérique, les plaques de Peyer et le sang périphérique. L'exploration du contrôle génétique *cis* de l'expression par l'identification d'expression allèle-spécifique a mis en évidence qu'environ 3 000 gènes étaient soumis à ASE dans les GALT et 1 200 dans le sang. Chaque tissu a montré des gènes avec un signal ASE spécifique, notamment 500 gènes et 592 gènes respectivement pour les plaques de Peyer jéjunale et iléale. Une liste de sept ASE a été soumise à des études de validation par Pyroséquençage.



C. Conclusions

Cette étude a permis pour la première fois d'étudier le profil transcriptomique complet des plaques de Peyer iléale et jéjunale chez le porc. Nos analyses ont ainsi révélé une prédominance de la différenciation des lymphocytes T vers les voies Th1, Th2, Th17 et iTreg dans la plaque de Peyer jéjunale, et la différenciation vers la voie Tfh dans la plaque de Peyer iléale. Ces résultats sont compatibles avec des études publiées qui ont montré une plus grande abondance de lymphocytes T que de lymphocytes B dans la plaque de Peyer jéjunale chez le porc.

Cette étude a également permis de montrer que l'expression des gènes dans le sang est fortement corrélée à l'expression de gènes dans des tissus lymphoïdes ce qui confirme l'intérêt de l'utilisation du sang comme tissu de remplacement pour l'étude de l'expression des gènes de tissus immunitaires spécialisés.



Immunome differences between porcine ileal and jejunal Peyer's patches revealed by global transcriptome sequencing of gut-associated lymphoid tissues

Tatiana Maroilley, Mustapha Berri, Gaëtan Lemonnier, Diane Esquerré, Claire Chevaleyre, Sandrine Mélo, Jean-Luc Coville, Jean-Jacques Leplat, Andrea Rau, Bertrand Bed'hom, Silvia Vincent-Naulleau, Marie-José Mercat, Yvon Billon, Patricia Lepage, Claire Rogel-Gaillard and Jordi Estellé

Abstract

The epithelium of the intestinal mucosa and the gut-associated lymphoid tissue (GALT) constitute an essential immunological and physical barrier against pathogens. In order to study the specificities of the GALT immunome in pigs, we compared the transcriptome profiles of jejunal and ileal Peyer's patches (PPs), mesenteric lymph nodes (MLN) and peripheral blood of four male piglets by RNA-Seq. We identified 1,103 differentially expressed (DE) genes between ileal and jejunal PPs, and six times more DE genes between PPs and MLN. The master regulators *FOXP3*, *GATA3*, *STAT4*, *TBX21* (which encodes T-bet), and *RORC* (which encodes Ror γ t) were down-regulated in ileal PPs compared to jejunal PPs, whereas the transcription factor *BCL6* was found to be overexpressed in ileal PPs. Our analyses revealed a predominance differential expressed genes between ileal and jejunal PPs related to the differentiation of T cell into Th1, Th2, Th17 and iTreg. Our results was also consistent with previous reports on a greater ratio of T cells on B cells in jejunal PPs compared to ileal PPs. We found antisense transcription for 24%, 22% and 14% of the transcripts detected in MLN, PPs and blood. Allele-specific expression analyses revealed both globally shared and tissue-specific cis-genetic control of gene expression.



Introduction

The digestive tract provides nutrients and energy from food to the organism and it is also a major potential entry site for pathogens. The epithelium of the intestinal mucosa, together with its associated immune system, are in direct contact with both symbiotic and pathogenic microorganisms and thus constitute an essential immunological and physical barrier¹. The specialized intestine immune system comprises the gut-associated lymphoid tissues (GALT) that include the Peyer's patches (PPs), the mesenteric lymph nodes (MLNs) and the isolated lymphoid follicles. PPs and MLNs both act as secondary lymphoid organs specifically associated with the gut tissues². MLNs are induction sites of immune responses, as they initiate peptide presentation to lymphocytes by antigen presenting cells arriving from the intestine². The PPs are distributed along the ileum and jejunum segments and are thus classified as either ileal PPs (IPPs) or jejunal PPs (JPPs). The PPs are covered with follicle-associated epithelium that harbors microfold cells or M cells. The sub-epithelial dome region contains abundant professional antigen presenting cells including dendritic cells (DCs) that take up antigens retrieved by M cells from the lumen of the digestive tract. The DCs present peptides to the T and B cells present in the PPs and trigger the initiation of antigen-specific immune responses and adaptive immunity. The PPs are also the primary inductive sites where most IgA immune responses are initiated by the class-switching of B cells from IgM to IgA antibodies³.

In humans, JPPs and IPPs are multiple isolated follicles, but in other species such as ruminants and pigs JPPs are isolated follicles whereas IPPs are instead continuous structures along the ileum⁴. The overall anatomical differences between JPPs and IPPs in pigs presuppose functional differences and their specific roles still need to be clarified^{5,6}. Indeed, several studies have described histological and functional differences between the two porcine Peyer's patches. For instance, the cellular composition of the GALT has been described as follows⁷ : B cells represent 90% and 45% of IPP and JPP cells, respectively, and 35% of MLN cells. Others studies have explored the functional and transcriptomic disparities for limited gene sets. For instance, it has been reported that IPPs would not be necessary for the development of the B cell systemic pool⁸ or specific IgA production. In contrast, in a targeted gene expression approach, Levast *et al.* (2010)⁹ showed differential



expression of 29 immune genes between IPPs and JPPs and observed a more diverse IgA repertoire in IPPs. Similarly, Gourbeyre *et al.* (2015)¹⁰ focused on pattern recognition receptors involved in innate immunity and observed differential expression for *TLR2*, *6*, *7*, *9*, *10* and *NALP3* between IPPs and JPPs.

The existing published results strongly suggest shared but subtle differences between the functionalities of IPPs and JPPs. To further explore this question, in this work we present a study of the transcriptome profiles of IPPs, JPPs, MLNs and peripheral blood in pigs with no clinical signs of infection by stranded RNA-sequencing. We report the differential gene expression between PPs and MLNs, and more specifically between IPPs and JPPs, together with the analysis of antisense transcription and allele-specific expression (ASE) in the four tissues.

Material and methods

Animals and sample collection

All animals were Large White pigs bred in the INRA experimental farm at Le Magneraud (GENESI, UE 1372, France). They belong to the Sus_Flora cohort previously reported¹¹. Thirty-six pigs were included in the present study. At 60 days of age, peripheral blood (PB) was sampled in the jugular vein using PAXgene Blood RNA tubes (PreAnalytiX, Qiagen, Germany). Following euthanasia at 70 days-of-age, MLN, IPPs and JPPs were collected as previously reported¹⁰, snap frozen in liquid nitrogen, and stored at -80°C until use. A subset of four uncastrated males was selected for the transcriptome study. The remaining group of 32 pigs that comprised 16 males and 16 females was used as validation group for RT-qPCR studies. Animals were weaned at 28 days of age and fed *ad libitum*. All animal experiments were carried out in accordance with European Guidelines for the Care and Use of Animals for Research Purposes. The animal protocol was approved by the local ethics committee in Poitou Charentes and assigned the approval number CE2013-2.

RNA extraction and sequencing



Tissue RNA extractions were performed as previously reported^{10,11}. In brief, 60 mg of tissue were lysed in 1 ml of Trizol (Invitrogen, Cergy Pontois, France) with ceramic beads (Bertin technologies, St Quentin en Yvelines, France), and total RNA was purified using RNeasy Mini Kit (Qiagen, Courtaboeuf, France) according to the manufacturer's recommendations. Residual genomic DNA was removed using DNase digestion with RNase-free DNase I Amplification Grade (Invitrogen, Cergy Pontoise, France) following the recommended protocol. For blood samples, no globin depletion procedure was applied, and RNA purification was performed as reported in Maroilley *et al.* (under revision). One µg of total RNA was reverse transcribed for 90 min at 37°C in a 20 µl volume containing 0.25 mM dNTP of OligodT, 25 U of MuMLV reverse transcriptase in 4 ml 5X MuMLV buffer (Eurogentec, Liège, Belgium). After heat-inactivation at 93°C for 5 min, generated cDNA was stored at -80°C until use. Finally, reverse-transcribed RNA samples of MLN, IPP, JPP and PB from the four selected pigs chosen were sequenced using 3 lanes in an Illumina HiSeq 2000 with the Illumina TruSeq Stranded mRNA Library Prep Kit. On average, 47 million reads (100-bp paired-end) were obtained per sample. The raw reads are available at NCBI's SRA repository (Bioproject PRJNA416857; accessions SAMN07966348 to SAMN07966363)

Differential expression analyses of RNA-Seq data

After quality control using the FastQC tool, sequencing reads were filtered with the Trimmomatic tool (v.0.32)¹² by trimming leading and trailing bases with phred qualities less than 3 and dropping reads shorter than 36 bases long and those with average phred qualities per base less than 15. Reads were subsequently aligned on the reference genome SusScrofa10.2 using TopHat2 (v2.0.14), with the default parameters and the gene annotation available for *Sus scrofa* in the Ensembl 85 release database¹³. Based on the read mapping with TopHat2, gene expression was quantified by obtaining read counts with the HTSeq-count software (v.0.6.1p1)¹⁴. Expression levels of antisense transcripts were also analyzed by using the strand read information. Read counts were further analyzed using the edgeR R/Bioconductor package (v.3.12.1)¹⁵. We retained genes as expressed in a tissue when the count per million (CPM) sense reads was greater than one for at least two animals. Similarly, a gene was found to have an antisense transcription if the antisense read



CPM was greater than one for at least two animals. Reads counts were then normalized according to the total number of reads of each sample using the Trimmed Mean of M-values normalization method¹⁶ implemented in edgeR (Supplementary Tables S1 and S2). The sense and antisense expression data structures were explored with multi-dimensional scaling (MDS) plots. The correlation between sense and antisense transcription across tissues was estimated by Spearman correlation using the normalized read counts.

The differentially expressed (DE) genes between GALT tissues were detected by fitting a negative binomial generalized linear model (GLM) in edgeR. In order to take into account intra-individual variability, the model included covariates for both tissue and individual. A likelihood ratio test was performed to identify DE genes among each pair of tissues, and P-values were corrected for multiple testing using the Benjamini-Hochberg control of the False Discovery Rate (FDR < 0.05). Finally, smear plots of log-fold expression changes versus log-concentration were produced using the “plotSmear” function in the edgeR package.

Joint genotyping and inference of allelic specific expression

For ASE analysis, we realigned reads along the reference genome SusScrofa10.2 using the STAR 2-pass protocol (v.2.4.0i)¹⁷. After marking duplicates, we split the reads, realigned them on indels and recalibrated the sequence data following the GATK (v3.7) Best Practices for calling variants in RNA-Seq data¹⁸, and as described in Maroilley *et al.* (under revision). SNPs were called with the HaplotypeCaller of the GATK tool (v3.7) using all samples and setting a base quality score > 10 and a coverage of >10 reads. We then removed SNPs with a low minor allele count (< 3 reads that harbor the alternative allele) using vcftools (v0.1.12a)¹⁹. In order to limit alignment biases of reads due to the presence of alternative SNP alleles, we created a masked genome reference sequence with an N at each SNP position using the Bedtools maskfasta tool (v2-2.24.0) and re-aligned reads on the masked genome with the same STAR 2-pass protocol. Using the dataset from this second alignment, counts of reads covering each allele at selected SNPs were obtained using ASEReadCounter of GATK v3.7 with parameters ensuring an adequate coverage and quality (read depth >10, mapping quality >10, base quality > 2). We then adapted ASEReadCounter output for the format required by QuASAR v.0.1²⁰ and used the associated statistical method to genotype individuals from



multiple RNA-Seq samples (fitAseNullMulti) and determine for each animal the probability of each possible genotype according to the alleles found in each tissue (with a minimum read coverage equal to five). Finally, we performed binomial tests on allelic read counts to detect potential allelic imbalance for the SNPs found to be heterozygous in each pig. The calculated P-values were corrected for multiple testing using the Benjamini-Hochberg control of the False Discovery Rate (FDR < 0.05).

Functional enrichment analysis

The *Sus scrofa* Ensembl gene IDs annotated from DE and ASE analyses were submitted to Biomart using the Ensembl 89 release in order to identify lists of human ortholog genes for subsequent functional enrichment analyses. The GOrilla tool²¹ was used for enrichment analyses of expressed genes in each tissue. For DE genes, we used the Ingenuity Pathway Analysis tool (IPA ; « <http://www.ingenuity.com> ») and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapper. SNPs and ASE SNPs were annotated with the Variant Effect Predictor tool (Ensembl).

Real-time quantitative PCR assay

Real-time quantitative PCR (RT-qPCR) was carried out on IPPs, JPPs and MLN from a group of 70-day-old pigs in order to validate the differential expression of a set of immune response mediator genes. The three genes *β-2-microglobulin (B2M)*, *hydroxymethylbilane synthase (HMBS)* and *glyceraldehyde-3-phosphate dehydrogenase (GAPDH)* were used as housekeeping reference genes because of their stable expression between samples as evaluated using the geNorm algorithm²². All gene primers were designed using the Clone Manager 9 software (Supplementary Table S3). For cDNA synthesis, one µg of total RNA was reverse transcribed for 90 min at 37°C in a 20 ml volume containing 0.25 mM dNTP of OligodT, 25 U of MuMLV reverse transcriptase in 4 ml 5X MuMLV biffer (Eurogentec, Liège, Belgium). After heat-inactivation at 93°C for 5 min, generated cDNA was stored at -80°C until use. RT-qPCRs were performed on a Bio-Rad Chromo 4 apparatus (Bio-Rad, Hercules, CA, USA) using the Mesa Green qPCR MasterMix Plus for SYBR assay (Eurogentec, Liège, Belgium) as previously described (Meurens *et al.* 2007; Levast *et al.* 2013). All



RT-qPCRs displayed efficiency between 90 and 110%, and expression data were given as relative values using delta delta Ct method after Genex macroanalysis (Bio-Rad, Hercules, CA, USA)²².

Pyrosequencing validation of selected ASE signals

A subset of ASE SNPs linked to six genes were selected for validation studies by pyrosequencing in a PyroMark Q24 instrument (Qiagen). Supplementary Table S4 shows primer sequences used for PCR amplification and subsequent sequencing. For each RNA sample, PCRs were performed using the HotStart Taq polymerase following the conditions specified in Supplementary Table S4. The allele quantification was subsequently performed in the PyroMark Q24 system according to manufacturer's instructions.

Results

Global transcription in GALT and peripheral blood

The transcriptome analyses were carried out by RNA-Seq on the same four piglets for blood samples (60 days of age), and for MLN, IPPs and JPPs (70 days of age). The global expression data structure was explored with a multi-dimensional scaling plot (MDS plot, Figure 1A) based on the TMM-normalized gene expressions for each sample. In this two-dimensional plot, samples are positioned according to the statistical distance of their expression profiles, highlighting here that samples are grouped by tissue and not by individual. Similar results were obtained from normalized antisense read counts (Figure 1B). For both sense and antisense transcription levels, we observed a clear separation between MLN, PB and PPs, whereas IPPs and JPPs grouped together, thus revealing relative similarities between both PP types.

Globally, we found over 13,000 expressed genes (according to the pig genome annotation) in IPPs, JPPs and MLNs. In PB, only ~9,200 genes were detected (Table 1), which is likely an underestimation since no globin depletion procedure was applied on blood samples, thus limiting the detection of weakly expressed genes due to an over-representation of globin genes. Indeed, HTSeq-count counted over 18 million reads for the



HBB (ENSSSCG00000014725) gene and almost 12 million reads for the *HBA* (ENSSSCG00000007978) gene corresponding to 43% and 27% of reads, respectively. Additionally, an HBB-like gene (ENSSSCG00000014727) was represented by four million reads (4% of the reads).

Out of the 14,596 genes found expressed in at least one tissue, 8,990 genes were co-expressed across the four tissues (Figure 1C). We found 3,632 genes shared by GALT but not expressed in PB. JPPs, IPPs and MLNs shared the same five most highly expressed genes, *cytochrome c oxidase subunit 1* (*COX1*), *cytochrome c oxidase subunit 2* (*COX2*), *cytochrome c oxidase subunit 3* (*COX3*), *ATP synthase F0 subunit 6* (*ATP6*) and *EEF1A1*. The *COX1*, *COX-2* and *COX3* mitochondrial genes, coding for cytochrome c oxidases, were represented by 8% and 5% of the sequencing reads in PPs and MLN, respectively. *ATP6* is also a mitochondrial gene coding for an ATP synthase, and *EEF1A1* codes for a translation elongation factor. The ten biological processes most enriched in genes expressed in all four tissues were functions related to energy transport and metabolic reactions (see Supplementary Table S5).

We quantified the transcription level of elements transcribed from the opposite strand of annotated genes in Ensembl (release 85) and found over 3,100 genes with measurable antisense transcription levels in PPs and MLN and less than half that number in PB (Table 1). Globally, 4,138 loci annotated as genes on one strand in Ensembl were detected as transcribed on the other strand in at least one tissue. The four tissues shared 1,258 genes with an antisense transcription. GALT samples shared 2,596 genes with an antisense expression and PPs shared 2,841 genes with a measurable antisense expression (Figure 1D). As presented in Table 1, the GALT and PB showed an antisense transcription for over 20% and 12% of expressed genes, respectively. The five genes with the largest number of antisense overlapping reads were the same in all tissues: *GYS1* (ENSSSCG00000003154) coding for glycogen (with 48% of antisense reads in blood), *NDUFB10* (ENSSSCG00000030328), *RNF167* (ENSSSCG00000017906), *ND6* (ENSSSCG00000018092) and *ENSSSCG00000012969*. Of note, we detected an antisense transcription level for 598 non-expressed genes in our dataset (see full list in Supplementary Table S6). Interestingly, GALT shared 230 common annotated genes harboring only antisense transcription, and 101 antisense expressed events were shared by the four tissues.



We mapped the sense and antisense reads along the genome (Supplementary Figure S1). We observed that they are not homogeneously distributed and that antisense transcription was preferentially localized on genomic regions showing sense transcription. Our results highlighted similar genomic regions across the four tissues that appear to be actively transcribed on both DNA strands.

We estimated correlations between sense and antisense gene transcription within and between tissues for the 14,596 genes with a CPM > 1 in at least one tissue and for the 4,138 genes with an antisense transcription level (CPM > 1 and > 10 reads). The correlation of sense and antisense gene expression levels was found to be higher between IPPs and JPPs than between PPs and MLN (Table 2). The correlation was lower between GALT and peripheral blood, and MLN appeared to be more correlated with PB. No significant correlations were found between the antisense and sense read counts for a same gene, whatever the tissue.

Differential gene expression analysis between GALT

We performed a differential expression analysis within the GALT to explore differences in the transcriptome between PPs and MLNs as well as between IPPs and JPPs, based on the 14,502 genes found to be expressed (sense transcription with CPM > 1) in at least one of the three tissues (Figure 2A). In a first exploratory approach, an MDS plot for the three GALT tissues (Figure 2B) showed that IPP and JPP samples were separated by intestinal section, revealing differences in gene expression between IPPs and JPPs. We subsequently tested whether gene expression was significantly different between the GALT tissues using the edgeR package (Table 3). The complete lists of DE genes are available in Supplementary Tables S7 to S9. There were more differences between MLN and PPs than between IPP and JPP. However, JPP seemed to be more similar to MLN than IPP. Smear plots showed that the bulk of genes was centered at a log-fold change of 0 indicating that any composition bias between libraries was successfully removed (Figure 2C-E). We observed that the scale of fold changes between MLN and PPs was larger than between IPPs and JPPs.



DE genes between the PPs and MLN were involved in signal transduction, immune response, and cell adhesion. They also affected metabolic processes such as lipid or organic metabolism processes (Figure 3). The DE genes between IPPs and JPPs were found to be involved in 60 pathways referenced in the KEGG database, where each of these pathways contained at least 10 DE genes (Supplementary Table S10). The pathway that included the most DE genes (103) was “Metabolic pathway”, and the second (40 DE genes) was “Cytokine-cytokine receptor interaction”. The pathway “Th1 and Th2 cell differentiation” was also enriched and included 19 DE genes, of which 18 and one were down- and up-regulated in IPP according to JPP, respectively (Figure 4).

By comparing these results to our previous study that targeted transcriptome analysis along the small intestine (duodenum, jejunum, ileum)¹¹, we found an overlap of 431 genes between the set of DE genes previously identified between jejunum and ileum and the 1105 DE genes between IPPs and JPPs. The fold change ranges were quite similar in both studies (respectively 1.7 and 1.9 log₂FC), and 97% of the shared genes showed a similar trend of differential expression, with only 14 genes presenting contradictory differential expression patterns between ileum/jejunum and IPP/JPP. The 431 shared DE genes were involved in pathways referred to as “LPS/IL-1 Mediated Inhibition of RXR Function”, “FXR/RXR Activation”, and “LXR/RXR Activation”. The top ten DE genes in term of absolute log-fold change most DE genes were assembled by IPA in various networks related to lipid metabolism (Supplementary Table S11), as reported in Mach *et al.* (2014)¹¹. In addition, over 60% of the DE genes (672 genes) between both PP types were not found DE in the intestinal segments in Mach *et al.*'s results. In enrichment analyses, DE genes specific to lymphoid tissues were found to be involved in several immunity-related canonical pathways: “Th1 and Th2 Activation Pathway”, “Primary Immunodeficiency Signaling”, “Th1 Pathway”, “iCOS-iCOSL Signaling in T Helper Cells”, and “T Cell Receptor Signaling”. The most enriched functions were “Hematological System Development and Function”, “Tissue Morphology”, and “Lymphoid Tissue Structure and Development”. The top ten DE genes between IPP and JPP assembled in IPA networks were related to biological functions linked to lymphoid tissue structure, cellular development and cell-mediated immune response.



Allelic specific expression analyses in porcine GALT and blood transcriptomes

Beyond the observed differential expression patterns between the GALT, we further explored the cis-genetic control of gene expression in each tissue using an ASE analysis. Overall, 432,198 SNPs were called across the 16 samples (MLN, IPP, JPP and PB from 4 individuals), after filtering. Each animal was genotyped according to the RNA-Seq data for the four tissues with the QuASAR R package²⁰, and 157,178 SNPs were found to be heterozygous in at least one animal (Supplementary Figure S2). While 46% of the SNPs were shared by at least two pigs and 5,600 SNPs were heterozygous for the four pigs, each animal still harbored a large number of specific heterozygous SNPs (17,224 to 25,160).

For each tissue, we tested SNPs heterozygous in at least one animal (Table 4) for possible ASE. In total, 157,178 heterozygous SNPs overlapping 12,755 genes were tested for ASE in at least one tissue. Results showed 16,797 significant ASE SNPs, 52% of which were found significant for every heterozygous animal (Table 4). Overall, 37% of the 12,755 genes tested had at least one SNP with ASE in at least one tissue (Table 4), and we found 2,509 genes affected by more than one heterozygous SNP ASE.

The proportion of genes showing significant ASE in at least one individual was close to 25 % for GALT and 17 % for blood. Of the genes showing significant ASE, 16% displayed ASE in the four tissues and were found to be involved in immune related functions such as antigen processing and presentation, T cell receptor signaling pathway of type I interferon signaling pathway by enrichment analyses. In blood, 141 genes were specifically under a cis-genetic control while they were also expressed in other tissues. In MLN, 669 genes specifically displayed an ASE. By comparing MLN and PPs, 113 genes expressed in GALT displayed specific ASE in PPs. 500 genes expressed in JPP and 572 genes expressed in IPP had ASE while expressed in the other tissues with no allelic imbalance (Supplementary Table S12). Nevertheless, the ASE genes specific to each of these tissues were not enriched in any functional category.



Validation of differentially expressed genes and ASE signals

RT-qPCRs were carried out for 24 genes (Supplementary Table S3) using tissues collected from the group of 32 pigs of 70 days of age belonging to the same cohort as the four animals included in the transcriptome analysis (Figure 5). The results confirmed that the *CCR10* and *CCR9* genes were under-expressed in IPP versus JPP. Furthermore, we showed that *CCR10* was less expressed in IPP compared to MLN, but more highly expressed in JPP compared to MLN and IPP. We also validated that *SLA-7* was not differentially expressed between GALT samples. RT-qPCR confirmed that *CCL28* and *IL15* were more expressed in PPs than in MLN, and that *SLA-6*, *SLA-8*, *MIC-2*, *MRI*, *IL10*, *CCR9*, *MADCAM1*, *NFKB1*, *SMAD2* and *SMAD3* were down-regulated in PPs in comparison to MLN.

As a first validation step of the ASE results, we compared the current ASE analysis on blood with a previous study that we conducted on 38 60-day-old Large-White pigs on peripheral blood (Maroilley *et al.*, in revision). Here, we revealed 1,005 ASE-SNPs (FDR < 0.05, at least 10 reads) found in at least two heterozygous pigs, of which 39 % of these SNPs (395) were already described as ASE-SNPs in our precedent study. Among them, 332 SNPs are described in dbSNP and include 55% missense variants, 43% synonymous variants and 1% stop gained variants. These 332 SNPs affected 93 genes.

In a second step, we used the pyrosequencing allele quantification approach to evaluate the consistency of the ASE effects detected in six genes and representing a range of allele frequencies and tissue specificities: *lysozyme (LYZ)*, *immunoglobulin lambda variable 10 (IGVL10)*, *complement factor D (CFD)*, *C-X-C motif chemokine ligand 13 (CXCL13)*, *2',-5',-oligoadenylate synthetase 2 (OAS2)* and *germinal center-associated, signaling and motility (GCSAM)*. Out of these six genes, we validated the RNA-Seq results in three of the cases representing a tissue-specific ASE (*LYZ*, with ASE confirmed only in MLN and PB) and two strong



ASE effects (*IGLV10* and *OAS2*). On the contrary, ASE effects of smaller magnitude could not be confirmed for *CFD*, *CXCL13* and *GCSAM*.

Discussion

Improving individual health in pig production systems requires a good understanding of immune and intestine pig functions to foster new disease preventive approaches. The pig species is also acknowledged as an excellent model for biomedical research, particularly for intestinal diseases due to anatomical and physiological similarities with the human digestive system²³. However, swine Peyer's patches (PPs) are an exception to this rule because, as in ruminants and rabbits, they have the particularity of forming a continuous structure in the ileum, while in humans they constitute independent follicles as in the pig jejunal and in all intestinal segments²⁴.

We report here a comparative analysis of the transcriptomic profiles of GALT in pigs including IPPs, JPPs and MLN, and PB as non-GALT reference immune tissue. The results showed extensive expression differences across GALT that were likely to reflect tissue specificities. The mapping of the *cis*- genetic control of gene expression with ASE analyses uncovered both shared and specific ASE signals in each tissue. Overall, our results propose a global picture of the transcriptomic control and profiles of GALT tissues in 70-day-old pigs in good health as a reference study to pave the way for a better understanding of the molecular mechanisms of Peyer's patches functions in pigs.

The porcine ileal and jejunal Peyer's patches have transcriptomic specificities linked with immunity

The anatomical differences between IPP and JPP in pigs have been previously described^{4,25}, and several studies have reported the transcriptome profiles of JPPs^{26,27}. However, to our knowledge, comparative analyses of gene expression between JPP and IPP in pigs have only been investigated in limited sets of target immune response-related-genes^{9,10}. Thus, our study is the first whole-transcriptome approach to this question, revealing 1,103 DE genes between IPPs and JPPs in pigs. The differential gene expression profiles between the two PP types could be partly related to their anatomical position along the small intestine. We



thus compared the IPP and JPP transcriptomes to those of the ileum and jejunum sections previously reported¹¹. We confirmed that genes found DE between IPPs and JPPs represented specific differences between the two PP types and were not simply reflecting the gene expression profiles of their corresponding intestinal sections.

We analyzed the gene sets found DE between IPP and JPP by performing IPA-based enrichment studies in order to explore which pathways were affected. We observed that canonical pathways related to T cell activity were down-regulated in IPP while the B cell receptor-signaling pathway was up-regulated in IPP compared to JPP. These differences were consistent with the variations in lymphocyte composition described in pigs, since IPP was reported to contain more B cells than JPP (90% and 45% of lymphocytes, respectively), and fewer T cells⁷. Several studies have described that the relative number of cells entering the IPPs was lower than that entering the JPPs in pigs^{28,29}. Accordingly, we found that the receptor *CCR7* and its ligand *CCL21* that mediate the organization of the T zone³⁰ were under-expressed in IPPs. The *CCR10* gene required for optimal T cell dependent IgA plasmocyte accumulation in mice³¹ was also found to be under-expressed in the IPP (logFC = -1.3). Furthermore, the receptor *CTLA4* involved in the regulation of activated T-lymphocytes³² was found to be under-expressed in the IPP, and *CCR9* promoting the homing of T cells towards the intestine³³ was found up-regulated in JPP (logFC = 0.8). Thus, all together our results are consistent with a higher proportion of T cells in the JPPs compared to the IPPs.

Our results highlighted gene expression differences between IPP and JPP related to the differentiation and fate of T cells (see Supplementary Table S10). The master regulators of CD4+ T cell commitment and plasticity have been reviewed³⁴. A slight but significant decrease in the master regulators known to be involved in naïve CD4+ T cell differentiation was observed in IPP compared to JPP (FC between -1.8 and -2.4, see Figure 4). The transcription factors found down-regulated in IPPs were FOXP3 (differentiation of induced T regulator cells or iTreg), GATA3 (Th2 differentiation), STAT4 and TBX21 that encodes T-bet (Th1 differentiation), as well as RORC that encodes Ror γ t (Th17 differentiation). Conversely, the transcription factor BCL6 was found overexpressed in IPP compared to JPPs (logFC=1.8). BCL6 is involved in the differentiation of the follicular helper T (Tfh) cells that have the important function of providing B-cell



help for the induction of antigen-specific antibody production³⁵. The Tfh cells contribute to the formation of germinal centers, which are dynamic microenvironments that provide a unique niche for B-cell affinity maturation to occur (review in Ramiscal and Vinuesa, 2013³⁶). ICOS, known to be expressed on surface of Tfh cells, was however found down-regulated in IPP (logFC = -1.4) as well as IL21 (logFC = -1.5), known to be produced by Tfh cells (Ma and Phan, 2017). It has been reported in mice that a fraction of Foxp3+ T cells can become Tfh cells in the PPs after down-regulating FOXP3 expression³⁷. This finding is in agreement with our results that revealed opposite differential expression between FOXP3 and BCL6 in IPP and JPP. Both JPP and IPP are main sites for priming B cells and educating the adaptive immune system while facing the gut microbiota. Tfh cells seem to play a crucial role in sensing microbiota-derived extracellular ATP and shaping commensal microbiota composition³⁸. Our results suggest that IPP is more abundant in Tfh and B cells than JPP, and this could be linked to the densification of the microbiota along the digestive tract. Even if we cannot exclude that the differential expression found between IPP and JPP could be solely due to distinct proportions of B and T cells, our results also suggested that T cells could have different fates in IPP and JPP.

Regarding the coherence of our results with previous reports, we observed that *CCR7-CCL21*, which mediate the organization of PPs into follicles, were down-regulated in the IPP. Moreover, we found that the *TNFSF11* gene involved in the development of GALT was down regulated in IPP in comparison with JPP (logFC < -1). Since the pigs in this study were 70 days old, this result is consistent with the described limited development of IPP that would be observed in adult pigs³⁹. In addition, Goubeyre *et al.* (2015)¹⁰ highlighted differential expression in JPPs and IPPs for TLR2, TLR6, TLR7, TLR9, TLR10, NALP3. We validated by RT-qPCR the differential expression of TLR7 (IPP < JPP) and TLR9 (IPP > JPP), and we identified that TLR8 was under-expressed in IPP compared to JPP. These correspondences indicate that the animal design employed in our study allows a description of the relevant differences between the two types of porcine Peyer's patches.

Major transcriptomic differences identified between MLN and PPs



MLN and PPs constitute the gut associated lymphoid tissues (with isolated lymphoid follicles) in swine. They participate in the activation of lymphocytes with the presentation of antigens passing through the intestine⁴⁰. Interestingly, TNFSF11, a gene involved in the development of lymph nodes and not of early steps in the PP development, was found under-expressed in PPs in comparison with MLN. Boeker *et al.* (1998)²⁵ explored the distribution of the major lymphocyte populations in the peripheral blood and lymphoid organs in six female Göttingen minipigs. They observed that the frequency of characterized cells in the Peyer's patches was lower than in other tissues. Our results were consistent with their histological data. Indeed, we observed an under-expression of the chemokines CXCL12, CXCL13, CXCR5, and CX3CR1 that mediate the attachment of cells to the lymphoid organs in PPs by comparison with MLN.

Antisense transcription profiles and ASE results are consistent with previous results

Beyond describing the differences in gene expression within GALT, we studied the regulation of gene expression by performing antisense transcript and ASE analyses in MLN, JPP, IMM and PB. We did not intend here to predict new antisense transcripts in intergenic regions, but only to quantify antisense transcription in annotated genes. Similarly to the work of He *et al.* (2008)⁴¹ on five different cell types, we observed that antisense transcription was not homogenous along the genome. Antisense transcription also seemed to be more frequent in regions with sense transcription. This might be due to a density reduction of the nucleosome structure where genes are transcribed, which may facilitate genome accessibility for antisense transcription. Chen *et al.* (2012)⁴² studied antisense transcription in liver and muscle of a pig F2 generation. They observed that 26% to 32% of liver transcripts and 46% to 54% muscle transcripts had an antisense transcription, respectively. In our study, we found 24%, 22% and 14% genes with antisense transcription in MLN, PPs and blood, respectively. The absence of significant correlations between sense and antisense transcription levels suggest that sense and antisense transcriptions are not co-regulated and are subjected to distinct regulation mechanisms.

Regarding ASE, we have found several effects across the four tissues (Table 4) and the enrichment analyses showed an overabundance of immune related functions associated with the genes shared between the four tissues. Overall, 37% of the tested genes displayed an ASE effect in at least one tissue, indicating the



extensive cis-regulation present in the immunome. A recent study conducted by Chamberlain *et al.*

(2015)⁴³ in cattle showed even more cis-regulation, with 89% of genes harboring heterozygous SNPs having at least one SNP with significant ASE. Similarly, Crowley *et al.* (2015)⁴⁴ reported that 89% of all genes tested in mouse brain show ASE.

We observed an important specificity in the cis-genetic control of gene expression between MLN and PPs, although intriguingly there were not specific enriched functions associated with these genes. For individual tissues, the proportion of genes showing significant ASE is close to 25% for GALT and 17% for blood, these proportions being in agreement with Chamberlain *et al.* (2015)⁴³ who previously reported percentages for white blood cells (33%), spleen (12%) and intestinal lymph (31%) in cattle. The GTEx consortium⁴⁵⁻⁴⁷ found that between 1.5 and 3.7% of 6,385 tested SNPs showed ASE; here we found globally 11% of tested SNPs showed an ASE, confirming a significant proportion of regulatory SNPs in the pig genome (Maroilley *et al.*, 2017, under revision)

The correlation of gene expression between blood and the GALT is consistent with use of blood as a source of biomarkers for the immune status assessment

Since the blood transcriptome is widely used as a source of biomarkers, we have performed global correlation analyses to evaluate whether it recapitulates at least in part the transcriptomes of GALT. The PPs, MLN and PB are complex multicellular tissues, and their gene expression profiles reflect the transcription levels of multiple cell types with an expected overrepresentation of genes expressed in the most abundant cell subset. As expected, transcripts from PB were dominated by the major blood proteins (hemoglobin and ferritine) and reduced the number of low expression genes that could be detected. Nevertheless, we observed strong positive correlations between the transcription levels of genes expressed in blood and GALT for both sense and antisense transcripts (Table 2). These data support the use of blood in pigs as a surrogate tissue and a potent source of biomarkers for functions taking place in other tissues. In addition, we found considerable overlap of ASE effects between GALT and blood, highlighting a shared cis-genetic control for the expression of immunity-related genes.



Figure legends

Figure 1. Summary plots of global gene expression analyses performed by RNA-Seq on MLN, IPP, JPP and PB. Graphs in the top show MDS plots for normalized gene expression levels (A) and normalized antisense transcription levels (B) and graphs on the bottom show Venn diagrams representing the number of expressed genes (C) and the number of genes with overlapping antisense reads (D), shared and specific to each tissue.

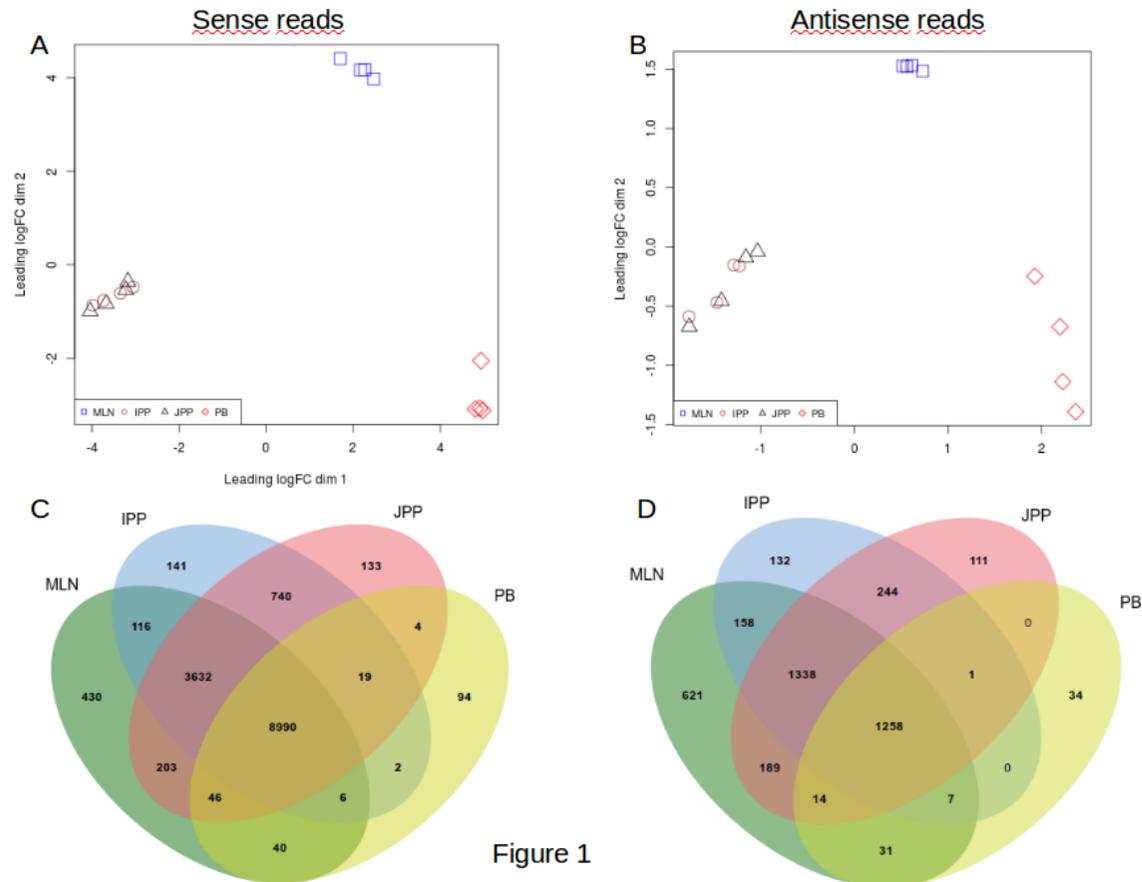


Figure 2. Summary plots of gene expression analyses within GALT tissues. Graphs show an MDS plot for normalized gene expression levels for GALT (A), a Venn diagram for the comparison of expressed genes in GALT (B) and smear plots of differential gene expression analyses when comparing IPP and JPP (C), MLN and IPP (D) and JPP and MLN (E).

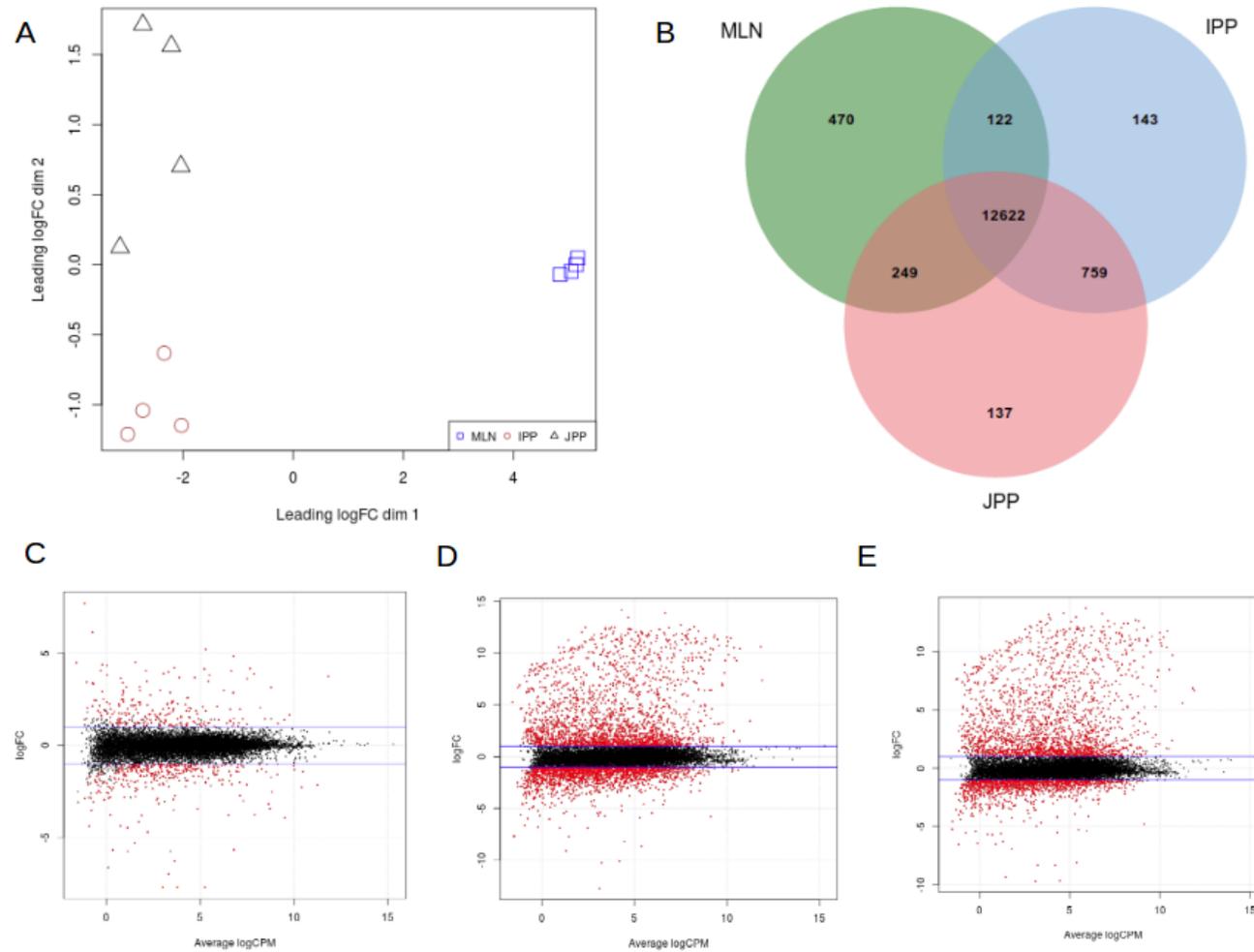


Figure 2

Figure 3. Ten most significantly enriched biological processes associated with differentially expressed genes between JPP and MLN (green) and IPP and MLN (in blue) performed with the GOrilla tool on the human ortholog GeneNames. The histograms represent the number of genes found DE and involved in a given biological process.

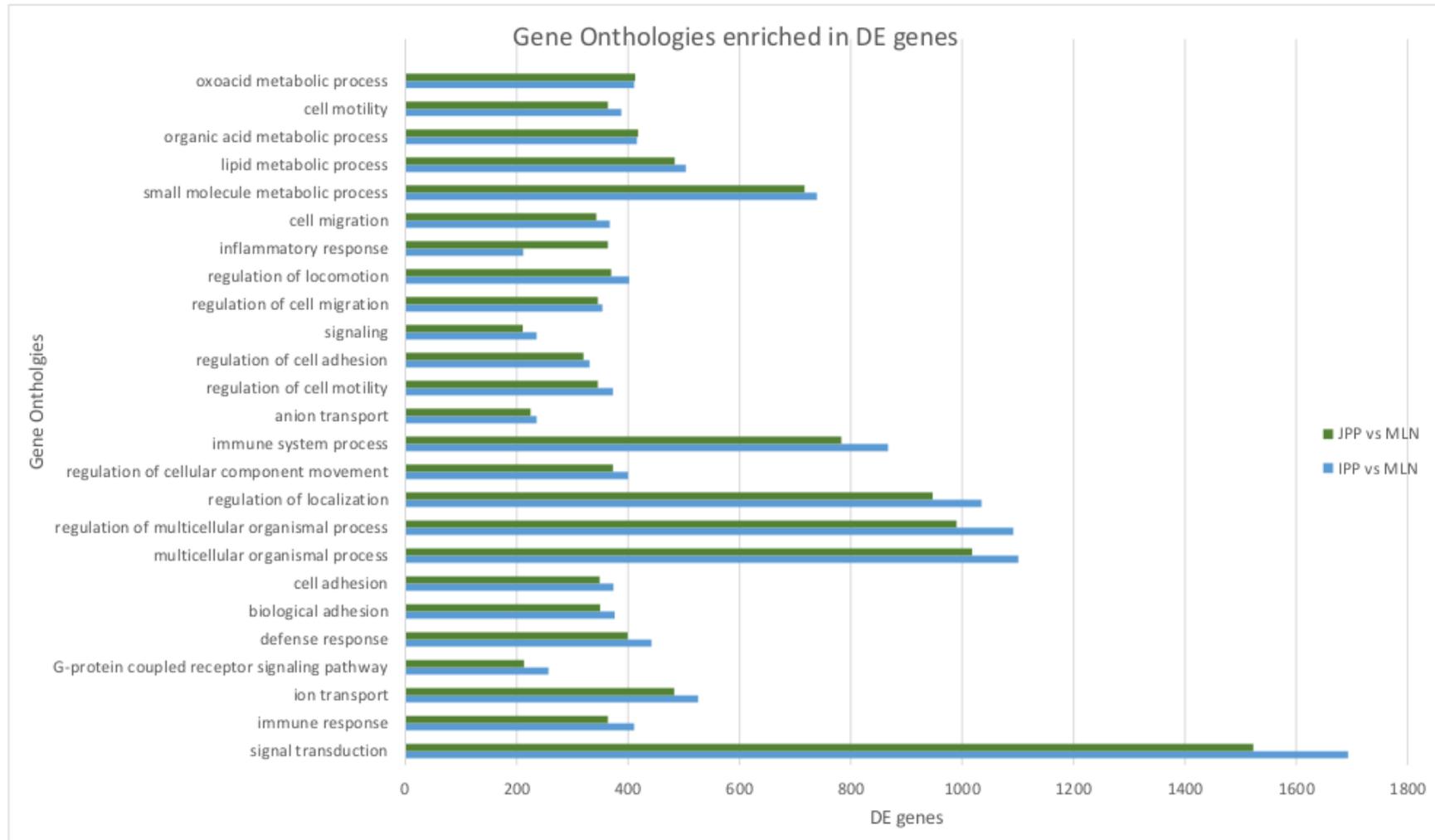
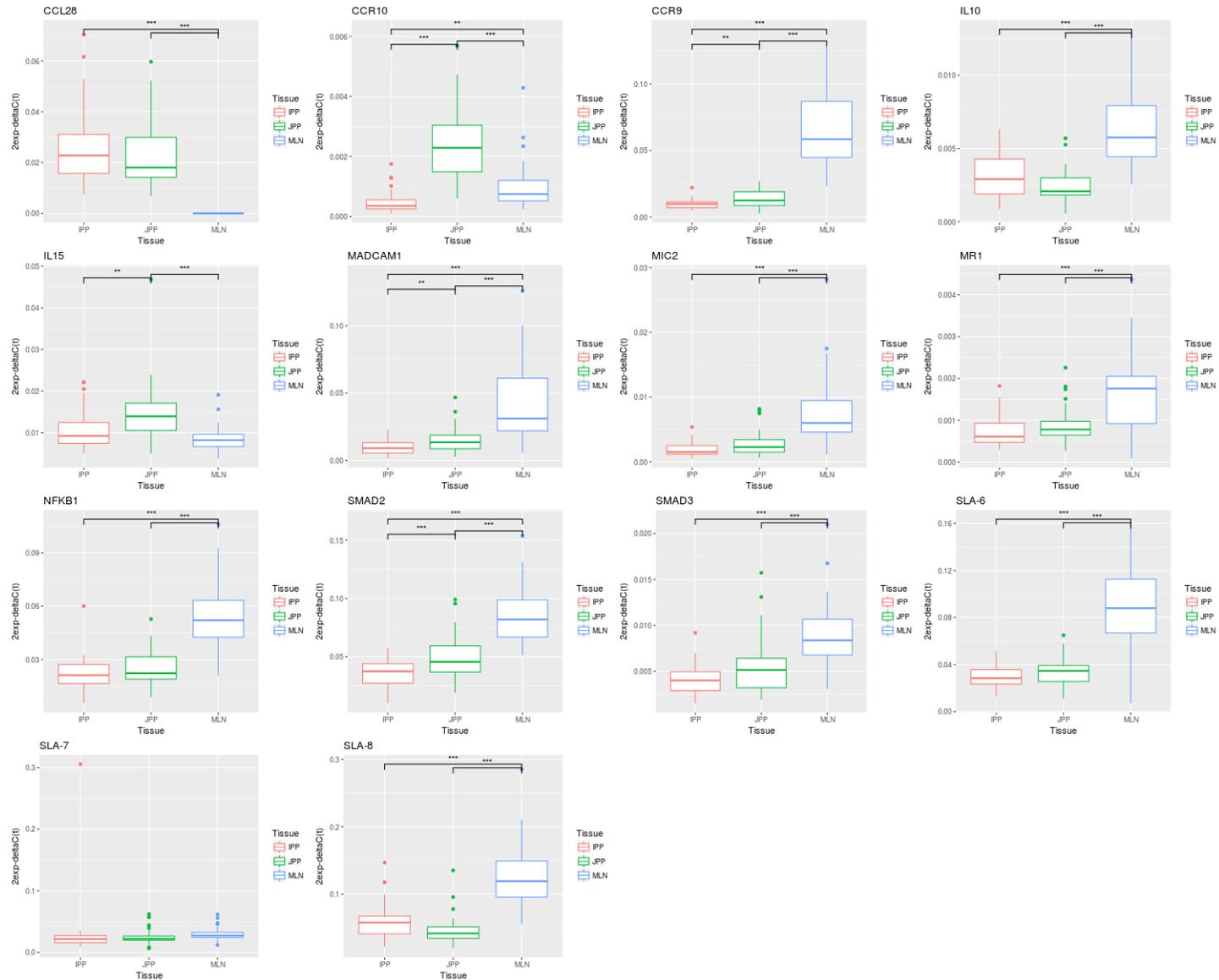


Figure 5. Real-time quantitative PCR analyses performed for the verification of differentially expressed genes found in RNA-Seq analyses within GALT tissues.



Tables

Table 1: Numbers of genes with sense and/or antisense transcription

	<i>JPP</i>	<i>IPP</i>	<i>MLN</i>	<i>PB</i>
<i>Expressed annotated genes</i>	13,767	13,646	13,463	9,201
<i>Number of genes that represented 50% of sense reads</i>	648	619	719	2
<i>Genes with antisense expression</i>	3,155	3,138	3,616	1,345
<i>Number of genes that represented 50% of antisense reads</i>	38	34	43	1
<i>Genes with both sense and antisense transcription</i>	2,841	2,836	3,242	1,059
<i>Genes harboring only antisense transcription</i>	(21%)	(21%)	(24%)	(12%)
	314	302	374	286

Table 2: Spearman correlation coefficients of gene expression (below diagonal), antisense expression (above diagonal) between tissues and in each tissue (diagonal).

Gene expression	Antisense transcription	MLN	IPP	JPP	PB
MLN		-0,04	0.72	0.72	0.71
IPP		0.83	-0,06	0.84	0.56
JPP		0.84	0.97	-0,06	0.58
PB		0.79	0.68	0.69	-0,1

Table 3. Differential gene expression among the GALT samples

	FDR < 0.05			FDR < 0.05 AND LOG₂FC >= 1		
	DE	Over-expressed	Under-expressed	DE	Over-expressed	Under-expressed
IPP VS MLN	6,852	3,405	3,447	3,922	2,143	1,779
JPP VS MLN	6,110	3,058	3,052	3,255	2,008	1,247
IPP VS JPP	1,103	444	659	549	217	332



Table 4. Summary of ASE results obtained on RNA-Seq datasets for the four tissues analyzed in this study (MLN, IPP, JPP and PB)

TISSUE	HETEROZYGOUS SNPS	GENES ¹	ASE SNPS (% TESTED)	CONSISTENT ASE SNPS ² (% ASE SNPS)	ASE GENES (% TESTED)	ASE GENES > 1 ASE SNP (% ASE GENES)
MLN	129,122	11,667	8,295 (6.4%)	4,046 (49%)	2,945 (25%)	1,543 (52%)
IPP	111,378	11,576	8,774 (7.9%)	4,265 (49%)	2,956 (26%)	1,524 (52%)
JPP	113,475	11,570	8,583 (7.5%)	4,161 (48%)	2,877 (25%)	1,420 (49%)
PB	40,584	7,410	3,339 (8.2%)	1,787 (54%)	1,254 (17%)	614 (49%)
TOTAL	157,178	12,755	16,787 (11%)	8,746 (52%)	4,769 (37%)	2,509 (53%)

¹Genes with a heterozygous SNP (< 5kb distance from gene according to Ensemble's VEP)

² For each animal heterozygote at this position, the SNP was detected with a significant ASE

Table 5. Allele quantification verification results performed with pyrosequencing on six genes having significant allelic specific expression SNP in RNA-Seq datasets.

		Animal 1			Animal 2			Animal 3			Animal 4		
		Genotype	RNAseq MAF	Pyroseq. cDNA MAF	Genotype	RNAseq MAF	Pyroseq. cDNA MAF	Genotype	RNAseq MAF	Pyroseq. cDNA MAF	Genotype	RNAseq MAF	Pyroseq. cDNA MAF
LYZ rs331048178	MLN	A/G	G: 16.5%	G: 24%	Homoz. A/A	-	-	Homoz. A/A	-	-	A/G	G: 21.1%	G: 23%
	IPP		A: 42.5%	A: 48%		-	-		A: 36.9%	A: 45%			
	JPP		n.s.	G: 46%		-	-		n.s.	G: 48%			
	PB		G: 34.3%	G: 34%		-	-		G: 26.9%	G: 34%			
IGLV10 rs710734656	MLN	A:75% T:25%	?	T: 24%	Homoz. A/A	-	-	A:70% T:30%	?	T: 15%	A:84% T:16%	?	T: 6%
	IPP		?	T: 10%		-	-		?	T: 5%			
	JPP		?	T: 12%		-	-		?	T: 5%			
	PB		?	T: 15%		-	-		?	T: 4%			
CFD rs326853901	MLN	Homoz. C/C	-	-	Homoz. C/C	-	-	C/T	n.s.	C: 49%	C/T	n.s.	C: 49%
	IPP		-	-		-	-		n.s.	C: 47%			
	JPP		-	-		-	-		C: 43.7%	C: 47%			
	PB		-	-		-	-		n.s.	A: 45%		n.s.	T: 45%
CXCL13 rs327682325	MLN	C/T	C: 44.7%	T: 49%	C/T	C: 39.7%	C: 48%	Homoz. A/A	-	-	Homoz. A/A	-	-
	IPP		n.s.	T: 47%		n.s.	T: 48%		-	-		-	-
	JPP		n.s.	T: 47%		n.s.	T: 49%		-	-		-	-
	PB		n.s.	-		n.s.	-		-	-		-	-
OAS2 rs80977910	MLN	Homoz.	-	?	C/G	C: 22.8%	C: 29%	Homoz.	-	?	C/G	C: 23.8%	C: 22%
	IPP		-	?		C: 24.8%	C: 21%		-	?		C: 29.3%	C: 19%
	JPP		-	?		C: 19.2%	C: 24%		-	?		C: 31.5%	C: 25%
	PB		-	?		C: 27.5%	C: 26%		-	?		C: 24.3%	c: 25%
GCSAM rs31 8341589	MLN	Homoz. G/G	-	-	G/T	n.s.	G: 50%	Homoz. G/G	-	-	G/T	n.s.	T: 48%
	IPP		-	-		T: 43.4%	G: 46%		-	-		T: 39.6%	T: 48%
	JPP		-	-		n.s.	G: 48%		-	-		n.s.	T: 48%
	PB		-	-		n.s.	G: 50%		-	-		n.s.	G: 50%



Supplementary Table S1: Count per million of reads mapping to annotated genes in the sense orientation for the four samples from the four pigs (Pig1 to Pig4)

Pig Genes Ensembl	Pig Gene Names ⁽¹⁾	Human Ortholog Gene Names ⁽¹⁾	Mesenteric lymph node (MLN)				Ileal Peyer's patches (IPPs)				Jejunal Peyer's patches (JPPs)				Peripheral blood (PB)			
			MLN	MLN	MLN	MLN	IPP	IPP	IPP	IPP	JPP	JPP	JPP	JPP	PB	PB	PB	PB
			PIG1	PIG2	PIG3	PIG4	PIG1	PIG2	PIG3	PIG4	PIG1	PIG2	PIG3	PIG4	PIG1	PIG2	PIG3	PIG4
ENSSSCG00000000001	NA	CELSR1	6.31	7.84	8.37	8.26	2.05	1.37	1.77	1.81	3.75	4.32	3.84	3.73	0	0.03	0	0.10
ENSSSCG00000000002	GTSE1	GTSE1	32.42	43.83	39.17	36.34	27.14	63.69	45.64	46.72	26.05	34.03	42.70	37.59	0.94	0.74	0.48	1.87
ENSSSCG00000000003	TTC38	TTC38	63.47	35.65	63.09	74.30	111.03	72.05	76.96	60.47	98.50	97.19	68.18	70.62	24.69	9.91	10.09	18.75
ENSSSCG00000000004	PKDREJ	PKDREJ	1.45	3.14	2.06	2.00	0.61	0.64	0.27	0.90	0.40	2.04	0.54	0.71	0	0	0	0
ENSSSCG00000000005	NA	CDPF1	8.95	7.72	11.04	11.07	4.36	4.80	8.02	7.13	6.19	5.16	9.61	7.35	5.60	2.63	1.83	3.74
ENSSSCG00000000006	PPARA	PPARA	0.28	0.17	0.09	0.24	1.01	0.60	1.95	3.14	4.97	2.01	2.07	5.62	0	0	0	0.03
ENSSSCG00000000007	TRMU	TRMU	32.95	27.73	32.60	31.59	23.42	22.64	25.65	20.79	26.05	26.64	25.85	21.18	8.94	4.51	4.04	6.91
ENSSSCG00000000010	NA	FBLN1	36.76	117.49	16.30	13.12	49.02	15.33	14.21	26.21	36.78	49.54	33.23	35.64	0.11	0.09	0.22	0.64
ENSSSCG00000000013	SMC1B	SMC1B	2.05	1.61	2.88	2.15	1.06	0.48	0.21	0.87	1.32	1.23	1.02	1.54	0	0	0	0
ENSSSCG00000000014	NA	NA	10.67	14.66	12.18	16.20	9.01	14.37	7.67	9.93	5.99	11.07	7.81	9.58	2.60	2.17	0.88	4.61
ENSSSCG00000000015	UPK3A	UPK3A	0.15	0.28	0.44	0.33	1.25	1.43	1.65	0.59	0.40	0.36	0.51	0.55	0.11	0.03	0.09	0
ENSSSCG00000000018	NA	KIAA0930	47.21	47.03	47.74	42.85	30.86	41.62	26.54	34.98	28.88	36.85	38.97	37.18	15.28	8.83	6.90	19.67
ENSSSCG00000000019	NA	NUP50	41.94	36.95	43.76	41.78	27.83	33.07	33.58	34.95	29.24	32.94	45.05	37.73	14.09	13.51	11.24	19.18
ENSSSCG00000000020	PHF21B	PHF21B	2.56	1.10	2.27	1.26	0.11	0.22	0.41	0.40	0.10	0.22	0.44	0.41	0	0	0	0.05
ENSSSCG00000000021	ARHGAP8	ARHGAP8	3.01	2.94	5.38	3.39	5.72	5.47	5.96	5.54	7.11	9.04	4.93	6.64	0.11	0.69	0.20	0.08
ENSSSCG00000000022	RTL6	RTL6	8.02	7.22	10.05	10.57	4.89	6.36	3.42	6.07	3.75	5.72	3.77	5.82	1.93	0.40	0.45	2.38
ENSSSCG00000000024	PARVB	PARVB	6.77	9.20	7.64	6.75	2.55	1.37	1.68	2.49	2.11	2.59	3.23	2.80	41.20	23.74	18.16	49.56
ENSSSCG00000000025	PARVG	PARVG	104.29	90.69	111.15	129.47	43.59	86.14	53.66	77.75	37.08	53.11	63.33	79.90	75.69	67.53	37.65	84.05
ENSSSCG00000000029	SCUBE1	SCUBE1	5.09	9.96	5.35	2.05	10.21	2.83	6.81	3.73	8.20	11.30	4.25	4.88	0.16	0	0.02	0.05
ENSSSCG00000000031	MCAT	MCAT	11.28	9.76	14.99	16.08	11.48	11.61	14.71	13.82	10.41	8.65	15.08	11.33	5.13	3.17	2.44	6.12
ENSSSCG00000000033	TSPO	TSPO	42.57	58.46	48.44	53.35	84.24	56.00	60.80	58.05	83.32	40.11	56.16	37.18	77.75	59.76	37.04	44.20



Supplementary Table S2: Count per million of reads mapping to annotated genes in the antisense orientation for the four samples from the four pigs (Fig1 to Fig4)

Pig Genes Ensembl	Pig Gene Names ⁽¹⁾	Human Ortholog Gene Names ⁽¹⁾	Mesenteric lymph node (MLN)				Ileal Peyer's patches (IPPs)				Jejunal Peyer's patches (JPPs)				Peripheral blood (PB)			
			MLN PIG1	MLN PIG2	MLN PIG3	MLN PIG4	IPP PIG1	IPP PIG2	IPP PIG3	IPP PIG4	JPP PIG1	JPP PIG2	JPP PIG3	JPP PIG4	PB PIG1	PB PIG2	PB PIG3	PB PIG4
ENSSSCG00000000001	NA	CELSR1	35.50	23.36	31.19	13.96	26.45	23.41	12.97	2.09	27.36	47.80	34.44	15.66	2.58	2.15	2.74	3.07
ENSSSCG00000000013	SMC1B	SMC1B	9.47	9.52	14.26	7.35	1.76	0.94	0	6.28	13.13	4.43	7.09	4.60	0	0	0	4.61
ENSSSCG00000000018	NA	KIAA0930	7.10	12.11	5.35	19.10	4.41	17.79	15.57	8.38	6.57	3.54	13.17	14.73	7.74	6.46	0	3.07
ENSSSCG00000000024	PARVB	PARVB	15.78	29.42	14.26	24.98	0	0.94	1.73	0	1.09	14.16	11.14	8.29	18.07	2.15	9.60	30.74
ENSSSCG00000000029	SCUBE1	SCUBE1	67.84	54.51	72.18	26.45	12.34	50.56	25.08	31.41	18.61	54.89	45.58	23.02	28.39	6.46	16.46	9.22
ENSSSCG00000000031	MCAT	MCAT	18.14	6.92	13.37	5.88	22.92	36.52	23.35	29.32	13.13	5.31	18.23	11.05	1.29	0	4.11	0
ENSSSCG00000000036	PACSIN2	PACSIN2	70.99	105.55	39.21	84.49	16.75	57.12	32.00	41.88	26.27	72.59	20.26	28.55	16.78	53.85	0	27.66
ENSSSCG00000000038	CYB5R3	CYB5R3	14.99	11.25	18.71	13.96	1.76	9.36	4.32	10.47	4.38	3.54	9.12	11.05	7.74	0	0	0
ENSSSCG00000000042	NA	NAGA	15.78	2.60	8.02	20.57	0	0	0.86	6.28	1.09	5.31	5.06	7.37	7.74	2.15	2.74	10.76
ENSSSCG00000000058	SNU13	SNU13	15.78	34.61	28.52	11.75	19.39	20.60	11.24	9.42	7.66	21.25	17.22	11.97	7.74	12.92	0	27.66
ENSSSCG00000000063	POLR3H	POLR3H	2306	2261	2221	2140	6163	5198	6489	5118	5905	5170	4318	4150	1099	872	570	1140
ENSSSCG00000000064	ACO2	ACO2	315.51	339.14	304.75	260.07	153.39	208.81	177.31	157.04	205.75	232.82	214.75	184.18	112.28	90.47	46.63	116.80
ENSSSCG00000000066	L3MBTL2	L3MBTL2	48.12	32.88	60.59	91.10	22.92	8.43	23.35	26.17	21.89	38.95	30.39	34.07	76.14	101.23	34.29	149.07
ENSSSCG00000000068	EP300	EP300	15.78	5.19	26.73	35.26	17.63	10.30	8.65	14.66	9.85	7.97	5.06	13.81	14.20	4.31	8.23	18.44
ENSSSCG00000000071	ST13	ST13	51.27	99.49	43.66	52.16	88.16	58.99	22.49	68.05	43.78	66.39	17.22	73.67	34.85	19.39	20.57	41.49
ENSSSCG00000000078	NA	TNRC6B	22.09	7.79	24.95	33.06	1.76	7.49	11.24	7.33	8.76	5.31	13.17	14.73	10.32	6.46	0	21.52
ENSSSCG00000000080	GRAP2	GRAP2	0	0	4.46	1.47	230.97	68.35	88.22	126.68	241.87	136.33	54.70	85.65	2.58	0	0	16.91
ENSSSCG00000000084	ATF4	ATF4	69.41	65.75	64.16	99.18	80.22	72.10	70.06	84.80	82.08	46.03	78.00	46.97	24.52	12.92	8.23	36.88
ENSSSCG00000000089	RPL3	RPL3	35.50	47.58	27.62	38.20	17.63	31.84	24.22	24.08	22.98	30.98	13.17	12.89	20.65	36.62	8.23	46.11
ENSSSCG00000000094	SUN2	SUN2	37.86	16.44	24.95	20.57	45.84	53.37	29.41	51.30	31.74	44.26	23.30	51.57	28.39	19.39	13.71	16.91
ENSSSCG00000000117	NA	C22orf23	87.56	48.45	20.50	41.88	54.66	65.55	27.68	21.99	48.16	78.79	20.26	21.18	33.56	23.69	9.60	13.83



Supplementary Table S3: Primer sequences used for RT-qPCR data

<i>Gene</i>	<i>Hybridation</i> <i>T°</i>	<i>Primer S</i>	<i>Primer AS</i>	<i>Amplicon size</i>	<i>NCBI n°</i>	<i>EST</i>
Alpha-4	60	GCTTTCAGAGCCAAATCC	ACCAGCCTTCCACATAACG	183	/	EW332552
April	60	TGCTCACCCGTAACAGAAG	TAAACTCCAGCATCCCAGAC	172	/	BP170456
B2MI	58	CAAGATAGTTAAGTGGGATCGAGAC	TGGTAACATCAATACGATTTCTGA	161	DQ845172	/
Baff	60	GAGAGCAGCTCCATTCAAAG	GCATGCCACTGTCTGCAATC	103	NM_001097498	/
CCL-25 bis	62	ACCTGCCTGCTGTGATATTC	TCCGATTGTCCAGGATCTTC	105	NM_001025214	/
CCL-28	64	CTGCTGCACTGAGTTTCAC	CAGCAGCCAAGTACAATCC	101	NM_001024695	/
CCR-09	69	TACGGCTATGACGCCACACC	ACGGCACCCACGATGAACAC	143	NM_001001624	/
GAPDH	63	CTCACGACCATGGAGAAGG	CCAAGCAGTTGGTGGTACAG	170	AF017079	/
HMBS2	58	AGGATGGGCAACTCTACCTG	GATGGTGGCCTGCATAGTCT	83	DQ845174	/
IL-08	62	TCCTGCTTTCTGCAGCTCTC	GGGTGGAAAGGTGTGGAATG	100	NM_213867	/
IL-10	60	GGTTGCCAAGCCTTGTGAC	AGGCACTTTCACCTCCTC	202	NM_214041	/
IL-15	60	CTGAGGATGGCATTTCATGTC	GGGATGAGCATCACTTTCAG	164	NM_214390	/
MAdCAM-1	68	AGCCTGGGCTCCGTAAGATC	TGGTCAGGGAAGGCCAACAC	155	NM_001037998	/
NFKB1	62	CGTGCTGCTGCTGCTGATG	CCGCCAAGGAGATGTTGTCG	111	DQ834921	/
SLA-6	60	ACTGATATAGC-AAAGAACA	AGAAAAGCTGTGCGACTCCGA	98	AY135602	
SLA-7	60	ACACGCATCTACAAGGACAC	TGGTAGGTGTGAGACCCGGC	98	AJ2519141	
SLA-8	60	ACACAGCGCGCAAGGATCT	TGGAAGGTATGAGACCCGGT	98	AJ2519141	
SLA-5	60	CATCATTGTTGGCCTGGTTC	CCTTTTTCACCTGAGCGC	89	AY247766, DQ104339, AY135602	
Smad-2	60	TGAGTGCTAAGTGACAGTG	CCAGAAGAGCAGCAAATTCC	143	/	DB812041
Smad-3	62	CGCAGAACGTCAACACCAAG	AGCTCATGGTGGCTGTGAAG	139	/	BX926114
SOCS2	58	CTGCGCATCGAATACCAAG	TGTAGAGCGGTTTGGTCAG	190	NM_001097461	/
STAT-3	60	TGCAGCAGAAAGTGAGCTAC	CCGGTCTTGATGACTAATGG	166	NM_001044580	/
TGFbR	58	AGTGGACTTGCCATCTTC	GTTCCCACTCTGTGGTTTG	194	AB182259	/
MIC2-3	64	CCTGGACCAGCTTACAGAG	ACCAGGCCACCGAGATATTC	131	NM_001114274.1	
MR1-3	60	TGGAGTACGGGAAAGATACC	CCATGAGCTCTGCAGTAAAG	105	XM_003130354	
CCR10	66	ctgttgatactgctgacactt	agaaggagacactgtgtctca	280	NM_001044563	
SLPI	62	AAGGTGAAGCCTGGGAAGTG	ACGTCTGGGAGTGGAATGG	187	NM_213870	



Supplementary Table S4: Primers and PCR conditions used for pyrosequencing

Name	Sequence 5'-3'	biotin
IGLV10_BFW	GATCGGGCTCCTCGCTGT	yes
IGLV10_RV	GCGGCTGTTTGTGCTGTAGA	
IGLV10_Seq	TGACTGACCCAGAGCTA	
GCNT3_BFW	CCTTGGGAAAAAAAAAACACTG	yes
GCNT3_RV	CTCATGGACCACCATAGAGTATGC	
GCNT3_S	GAGTATGCGATGGGG	
GCSAM_FW	ATCATCTTCTCCCTTCCAGG	
GCSAM_BRV	AAGATTCTGCTGGGCATGA	yes
GCSAM_S	TGTGTCTTCTACTCCTGG	
CFD_FW	CTGCAACCTGCGCACATAC	
CFD_BRV	GCGCTGTCAGCCATCACT	yes
CFD_S	CTACGTGGCCTGGAT	
OAS2_FW	AGGGAACCGCTGGTGTG	
OAS2_BRV	CCAGCACTATCAGTCCATGAATC	yes
OAS2_S	GGAAGCTGTGGAGCA	
IGLV4_FW	ATGTGGCCTCTTCTAAGCTGACT	
IGLV4_BRV	GCGGTATTTAGACCATTTACTC	yes
IGLV4_S	TCGCCTGCCAAGGAG	
LYZ_FW	AAGGTCTATGATCGGTGCGAGTTC	
LYZ_BRV	TGCAGGATATGTGACAGGCATTA	yes
LYZ_S	GCCGCTACTGGTGTAA	
CXCL13_BFW	CTTGAACCCTCAAAGCAAAGTGT	yes
CXCL13_RV	ATCTTCTGGAAAAGAGCCAATCA	
CXCL13_S	CGTGCAGGTGTCCTT	



Supplementary Table S5: The 10 most significant processes enriched in expressed genes in each tissue

MLN								
GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0090662	ATP hydrolysis coupled transmembrane transport	2.92E-45	4E-41	12.43	10844	49	801	45
GO:0099131	ATP hydrolysis coupled ion transmembrane transport	7.24E-42	4.95E-38	12.62	10844	44	801	41
GO:0099132	ATP hydrolysis coupled cation transmembrane transport	1.99E-36	9.06E-33	13.15	10844	35	801	34
GO:0006631	fatty acid metabolic process	9.34E-33	3.19E-29	15.22	10844	192	141	38
GO:0044281	small molecule metabolic process	1.16E-29	3.16E-26	3.09	10844	1137	383	124
GO:0032787	monocarboxylic acid metabolic process	3.15E-28	7.19E-25	10.30	10844	306	141	41
GO:0044255	cellular lipid metabolic process	8.74E-28	1.71E-24	4.02	10844	621	378	87
GO:0019752	carboxylic acid metabolic process	9.85E-28	1.68E-24	4.15	10844	573	383	84
GO:0006629	lipid metabolic process	5E-26	7.59E-23	3.56	10844	758	378	94
GO:0006082	organic acid metabolic process	5.77E-25	7.89E-22	3.73	10844	653	383	86

IPP								
GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0090662	ATP hydrolysis coupled transmembrane transport	1.02E-49	1.41E-45	12.53	11044	53	815	49
GO:0099131	ATP hydrolysis coupled ion transmembrane transport	3.21E-45	2.22E-41	12.69	11044	47	815	44
GO:0099132	ATP hydrolysis coupled cation transmembrane transport	1.09E-38	5.04E-35	13.18	11044	37	815	36
GO:0006631	fatty acid metabolic process	8.3E-35	2.87E-31	15.40	11044	202	142	40
GO:0032787	monocarboxylic acid metabolic process	4.78E-31	1.32E-27	10.59	11044	323	142	44
GO:0044281	small molecule metabolic process	1.81E-29	4.18E-26	3.01	11044	1197	393	128
GO:0051056	regulation of small GTPase mediated signal transduction	1.77E-28	3.49E-25	5.07	11044	242	621	69
GO:0044255	cellular lipid metabolic process	2.76E-27	4.78E-24	3.92	11044	661	375	88
GO:0019752	carboxylic acid metabolic process	4.18E-27	6.42E-24	4.08	11044	599	380	84
GO:0006082	organic acid metabolic process	1.09E-25	1.52E-22	6.10	11044	676	142	53

JPP								
GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0090662	ATP hydrolysis coupled transmembrane transport	1.3E-48	1.81E-44	12.52	11118	52	820	48
GO:0099131	ATP hydrolysis coupled ion transmembrane transport	4.12E-44	2.87E-40	12.67	11118	46	820	43



GO:0099132	ATP hydrolysis coupled cation transmembrane transport	1.08E-38	5E-35	13.19	11118	37	820	36
GO:0006631	fatty acid metabolic process	1.22E-33	4.25E-30	15.22	11118	202	141	39
GO:0032787	monocarboxylic acid metabolic process	7.59E-30	2.11E-26	10.37	11118	327	141	43
GO:0044281	small molecule metabolic process	1.2E-28	2.79E-25	2.97	11118	1203	395	127
GO:0051056	regulation of small GTPase mediated signal transduction	9.64E-28	1.92E-24	4.95	11118	248	625	69
GO:0019752	carboxylic acid metabolic process	1.74E-27	3.03E-24	4.08	11118	608	381	85
GO:0044255	cellular lipid metabolic process	2.66E-27	4.13E-24	3.92	11118	663	376	88
GO:0006082	organic acid metabolic process	1.11E-25	1.54E-22	6.09	11118	686	141	53

PB								
GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0090662	ATP hydrolysis coupled transmembrane transport	2.23E-37	2.76E-33	12.70	7893	38	589	36
GO:0099131	ATP hydrolysis coupled ion transmembrane transport	4.94E-34	3.06E-30	12.63	7893	35	589	33
GO:0099132	ATP hydrolysis coupled cation transmembrane transport	9.01E-31	3.71E-27	12.95	7893	30	589	29
GO:0006631	fatty acid metabolic process	5.04E-28	1.56E-24	16.41	7893	142	105	31
GO:0032787	monocarboxylic acid metabolic process	8.79E-24	2.17E-20	10.98	7893	226	105	33
GO:0044255	cellular lipid metabolic process	7.53E-23	1.55E-19	4.31	7893	460	263	66
GO:0019752	carboxylic acid metabolic process	4.56E-22	8.05E-19	4.31	7893	434	270	64
GO:0006082	organic acid metabolic process	1.34E-21	2.07E-18	6.50	7893	486	105	42
GO:0043436	oxoacid metabolic process	1.05E-20	1.44E-17	6.39	7893	482	105	41
GO:0044281	small molecule metabolic process	5.16E-20	6.38E-17	2.96	7893	875	274	90



Supplementary Table S6: Genes overlapped only with antisense reads

Pig_Ensembl_Genes	Pig_GeneNames	Human_GeneNames	MLN	IPP	JPP	PB
ENSSSCG00000000029	SCUBE1	SCUBE1				x
ENSSSCG00000000117		C22orf23		x	x	x
ENSSSCG00000000228	SCN8A	SCN8A	x			
ENSSSCG00000000326		OR6C75	x			
ENSSSCG00000000411		NAB2	x	x	x	x
ENSSSCG00000000443		GLI1				x
ENSSSCG00000000651	CLEC12B		x			x
ENSSSCG00000000655	KLRF1	KLRF1	x	x	x	x
ENSSSCG00000000709	PLEKHG6	PLEKHG6	x			x
ENSSSCG00000000749	SLC6A12	SLC6A12				x
ENSSSCG00000000807	SLC38A1	SLC38A1				x
ENSSSCG00000000852					x	
ENSSSCG00000000858	PMCH	PMCH	x	x	x	x
ENSSSCG00000000863	SYCP3	SYCP3	x	x	x	x
ENSSSCG00000000894	CCDC38	CCDC38	x	x	x	
ENSSSCG00000000940		PPFIA2		x	x	
ENSSSCG00000000941	METTL25	METTL25				x
ENSSSCG00000001001		C6orf201	x	x	x	x
ENSSSCG00000001027	BMP6	BMP6				x
ENSSSCG00000001088	ATP6V1G1	ATP6V1G1	x	x	x	
ENSSSCG00000001089	GPLD1	GPLD1	x	x	x	



Supplementary Table S7: Differentially expressed genes between ileal Peyer's patches and mesenteric lymph nodes

Pig Ensembl Genes	Pig GeneNames	Human GeneNames orthologs	Fold Change	log2FC	log2CPM	LR	PValue
ENSSSCG00000025267		SLC26A3	19024,8922	14,21560066	4,289662921	183,2274598	9,57E-42
ENSSSCG00000016135	CPO	CPO	14874,4001	13,86054386	6,652883133	90,63959988	1,72E-21
ENSSSCG00000027361	OVOL1	OVOL1	11087,20374	13,43660793	4,658946448	677,2824608	2,61E-149
ENSSSCG00000021971	DPEP1	DPEP1	6785,288908	12,72819453	7,499548666	292,9072341	1,16E-65
ENSSSCG00000017120	SLC6A19	SLC6A19	5962,685072	12,54174643	8,219605535	962,7116273	2,29E-211
ENSSSCG00000012855		SLC22A18	5754,422489	12,49045543	3,756444718	427,1785105	6,68E-95
ENSSSCG00000028309	CH242-64C23.1	REG3G	5727,428501	12,48367183	2,340007844	91,6557115	1,03E-21
ENSSSCG00000017763		SLC13A2	5697,645993	12,47615027	3,925811776	328,0546209	2,55E-73
ENSSSCG00000024743	PRR15L	PRR15L	5605,61987	12,4526582	3,427836911	443,9614281	1,49E-98
ENSSSCG00000009317	CDX2	CDX2	5581,733811	12,44649761	6,85141215	877,5992578	7,27E-193
ENSSSCG00000005861			5288,97002	12,36877108	3,639146192	375,892243	9,76E-84
ENSSSCG00000023367	IL22RA1	IL22RA1	5203,56994	12,34528602	4,678887198	378,2304572	3,02E-84
ENSSSCG00000011962		ADGRG7	5148,472833	12,32992884	4,691434005	705,0661031	2,37E-155
ENSSSCG00000016204	IHH	IHH	5105,893979	12,31794787	4,797299835	557,4135721	3,07E-123
ENSSSCG00000024336		TMPRSS2	5071,577597	12,30821888	6,339182986	1008,094287	3,13E-221
ENSSSCG00000016284		ALPP	5010,316293	12,29068597	7,984324681	835,4198759	1,07E-183
ENSSSCG00000022133		RAB17	4992,842387	12,28564565	3,42663027	514,8164072	5,68E-114
ENSSSCG00000006820	EPS8L3	EPS8L3	4763,924972	12,21793498	7,862088467	918,8689959	7,76E-202
ENSSSCG00000028330		APOB	4595,699587	12,16606878	5,885255701	563,3999486	1,53E-124
ENSSSCG00000001724	MEP1A	MEP1A	4458,282113	12,1222722	8,445048759	594,0527545	3,29E-131
ENSSSCG00000003390		SLC2A7	4333,980847	12,08147706	5,339655174	610,8511695	7,3E-135
ENSSSCG00000016309		UGT1A1	4317,988248	12,0761436	3,864604908	319,6185675	1,75E-71



Supplementary Table S8: Differentially expressed genes between jejunal Peyer's patches and mesenteric lymph nodes

Pig_Ensembl_Genes	Pig_GeneNames	Human GeneNames Orthologs	Fold Change	logFC	logCPM	LR	PValue
ENSSSCG00000010997	SPINK4	SPINK4	4547,150954	12,15074718	6,200426569	1203,730917	9,43E-264
ENSSSCG00000005979	ANXA13	ANXA13	1273,032323	10,31405334	6,984162992	1109,38861	3,01E-243
ENSSSCG00000027387	MS4A12	MS4A12	1509,608968	10,55995918	8,29150975	1101,435664	1,61E-241
ENSSSCG00000002825		CES1	705,1500499	9,461786473	6,441044221	1081,119322	4,19E-237
ENSSSCG00000000634	STYK1	STYK1	621,5210989	9,279659557	6,278254101	1035,275942	3,86E-227
ENSSSCG00000011703	TM4SF4	TM4SF4	1943,752787	10,92462903	8,03010098	1018,709666	1,54E-223
ENSSSCG00000028173	TRIM31	TRIM31	2725,771216	11,41244876	6,200274563	1003,43264	3,22E-220
ENSSSCG00000015241	TMEM45B	TMEM45B	1958,264641	10,93536003	6,692686857	1002,618153	4,84E-220
ENSSSCG00000015363	AGR2	AGR2	423,3160891	8,725591513	6,852468541	995,3606411	1,83E-218
ENSSSCG00000006712	REG4	REG4	3545,390358	11,79172877	6,883965257	980,1316443	3,74E-215
ENSSSCG00000013480	SMIM24	SMIM24	2648,314551	11,37085877	8,288120956	975,4291848	3,94E-214
ENSSSCG00000017120	SLC6A19	SLC6A19	5575,343384	12,44484495	8,219605535	951,8177281	5,34E-209
ENSSSCG00000007850	ANKS4B	ANKS4B	1143,982522	10,15984929	6,453600323	948,1865325	3,29E-208
ENSSSCG00000024336		TMPRSS2	3624,395601	11,82352472	6,339182986	938,116053	5,08E-206
ENSSSCG00000004703		CKMT1A	2271,123921	11,14919071	8,842039071	936,9841512	8,96E-206
ENSSSCG00000013425	MISP	MISP	2259,268969	11,14164032	7,247282255	924,2545185	5,24E-203
ENSSSCG00000030453			1768,784177	10,78854231	7,78819936	917,271379	1,73E-201
ENSSSCG00000006820	EPS8L3	EPS8L3	4314,223155	12,07488509	7,862088467	904,0041788	1,32E-198
ENSSSCG00000025885		ITLN2	2302,39384	11,16891892	5,3411817	900,5937252	7,29E-198
ENSSSCG00000007371	HNF4A	HNF4A	2981,604819	11,54187334	7,727117454	898,8760755	1,72E-197
ENSSSCG00000016247	TM4SF20	TM4SF20	3160,11386	11,62576082	7,767754914	896,2573847	6,39E-197
ENSSSCG00000014985	MMP1	MMP1	4544,400193	12,14987417	7,194532495	892,6744661	3,84E-196



Supplementary Table S9: Differentially expressed genes between ileal Peyer's patches and jejunal Peyer's patches

Pig_Ensembl_Genes	Pig_GeneName	Human GeneNames Orthologs	Fold Change	logFC	logCPM	LR	PValue
ENSSSCG00000026594	0	0	0,000266532	-11,87340338	3,762737799	468,0954063	8,32E-104
ENSSSCG00000010432	0	ASAH2	0,001920599	-9,024227653	5,238572713	208,8332078	2,47E-47
ENSSSCG00000006551	AQP10	AQP10	0,003096269	-8,335253347	2,996373389	136,6081552	1,47E-31
ENSSSCG00000022383	GATA4	GATA4	0,007996886	-6,966345981	3,30175814	191,9026304	1,22E-43
ENSSSCG00000001521	MLN	MLN	0,0101455	-6,623016195	0,08098888	82,74569388	9,33E-20
ENSSSCG00000028204	0	TREH	0,012960154	-6,269773323	3,621678544	43,54872318	4,14E-11
ENSSSCG00000023508	0	BCO1	0,018844571	-5,729707238	1,141637706	41,23321334	1,35E-10
ENSSSCG00000012019	TMPRSS15	TMPRSS15	0,018865405	-5,72811309	3,405972203	139,1143626	4,16E-32
ENSSSCG00000023939	0	TREH	0,019685523	-5,666721133	0,328619892	68,88896731	1,04E-16
ENSSSCG00000017163	ENPP7	ENPP7	0,019910888	-5,65029866	6,795599421	248,3510339	5,94E-56
ENSSSCG00000021490	PRDM16	PRDM16	0,025649102	-5,284947892	1,504415092	30,51899486	3,31E-08
ENSSSCG00000023612	ACE3	0	0,038690292	-4,691884577	2,190311618	41,79297971	1,01E-10
ENSSSCG00000011256	0	SLC22A13	0,045458155	-4,459317066	1,110927348	16,27566344	5,48E-05
ENSSSCG00000003845	0	MROH7	0,047987985	-4,381182942	-0,925822044	22,16863394	2,50E-06
ENSSSCG00000009312	PDX1	PDX1	0,048261307	-4,372989196	0,737647513	27,42921314	1,63E-07
ENSSSCG00000028896	0	DIO1	0,056394545	-4,148300578	1,718179638	76,29295056	2,45E-18
ENSSSCG00000030013	0	LCT	0,065626944	-3,929567942	9,506252639	55,69742096	8,45E-14
ENSSSCG00000017295	KCNH6	KCNH6	0,071650861	-3,802872152	1,872640881	38,43996299	5,65E-10
ENSSSCG00000015691	0	LCT	0,072414059	-3,787586369	6,50456393	59,62422384	1,15E-14
ENSSSCG00000011147	0	AKR1C2	0,073222202	-3,771575024	0,20562481	21,11322589	4,33E-06
ENSSSCG00000011663	RBP2	RBP2	0,074147303	-3,753461974	7,902368914	104,3431613	1,70E-24
ENSSSCG00000024627	0	0	0,087243687	-3,518805448	-0,806149126	24,19919369	8,69E-07



Supplementary Table S10: KEGG pathway mapping for DE genes between IPP and JPP

Metabolic pathways (103)									
ABAT	ACADSB	ACSL3	ACSM5	ACSS1	ADA	ADH1A	AGPAT3	AHCYL2	AKR1B1
Cytokine-cytokine receptor interaction (40)									
CCL17	CCL19	CCL20	CCL21	CCL22	CCR10	CCR5	CCR7	CCR9	CD27
PI3K-Akt signaling pathway (29)									
ANGPT2	BAD	BCL2L1	CCND2	CDK2	CDKN1A	CREB3L1	CREB3L2	CREB3L3	FGF14
Pathways in cancer (25)									
ADCY4	ADCY7	BAD	BCL2L1	BDKRB1	CDK2	CDKN1A	CKS1B	E2F1	FGF14
Chemokine signaling pathway (23)									
ADCY4	ADCY7	CCL17	CCL19	CCL20	CCL21	CCL22	CCR10	CCR5	CCR7
Th17 cell differentiation (23)									
AHR	CD247	CD3D	CD3E	CD3G	CD4	FOXP3	GATA3	HLA-DQB1	IFNGR1
HTLV-I infection (23)									
ADCY4	ADCY7	BCL2L1	BUB1B	CCNB2	CCND2	CD3D	CD3E	CD3G	CDC20
Human papillomavirus infection (20)									
BAD	BCL2L1	CCND2	CDK2	CDKN1A	CREB3L1	CREB3L2	CREB3L3	E2F1	ITGA2
T cell receptor signaling pathway (20)									
CD247	CD28	CD3D	CD3E	CD3G	CD4	CD40LG	CD8A	CD8B	CTLA4



Supplementary Table S11: DE genes between IPP and JPP related to IPA functional networks

Pig_Gene_stable_ID	GeneID	logFC	logCPM	LR	PValue	DE IL/JE	IPA networks
ENSSSCG00000026594		-11,87	3,76	468,10	8,32E-104	yes	
ENSSSCG00000009216	SPP1	4,82	6,76	338,60	1,29E-75	yes	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry
ENSSSCG00000017163	ENPP7	-5,65	6,80	248,35	5,94E-56	yes	Digestive System Development and Function, Organ Morphology, Organismal Development
ENSSSCG00000010432	ASAH2	-9,02	5,24	208,83	2,47E-47	yes	Lipid Metabolism, Small Molecule Biochemistry, Visual System Development and Function
ENSSSCG00000027867	CSTA	3,88	3,04	198,36	4,77E-45	yes	Drug Metabolism, Molecular Transport, Hereditary Disorder
ENSSSCG00000022383	GATA4	-6,97	3,30	191,90	1,22E-43	yes	Cardiovascular System Development and Function, Cellular Assembly and Organization, Cellular Development
ENSSSCG00000003908	TSPAN1	2,98	6,25	190,62	2,33E-43	yes	Cell-To-Cell Signaling and Interaction, Cellular Movement, Cardiovascular Disease
ENSSSCG00000025125	SLC10A2	5,18	5,29	174,69	7,01E-40	yes	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry
ENSSSCG00000010437	PAPSS2	-3,00	7,30	164,53	1,16E-37	yes	Drug Metabolism, Molecular Transport, Hereditary Disorder
ENSSSCG00000013073	FADS3	2,85	6,75	141,08	1,54E-32	yes	Embryonic Development, Hematological System Development and Function, Lymphoid Tissue Structure and Development



ENSSSCG00000010077	IGLL1	2,56	4,35	82,84	8,91E-20	no	Lymphoid Tissue Structure and Development, Tissue Morphology, Hematological System Development and Function
ENSSSCG00000030787		-2,11	6,16	73,77	8,76E-18	no	
ENSSSCG00000001639	TRERF1	-1,43	5,10	66,82	2,98E-16	no	Cellular Development, Cellular Growth and Proliferation, Connective Tissue Development and Function
ENSSSCG00000013594	FBN3	2,97	0,88	64,86	8,04E-16	no	Cancer, Organismal Injury and Abnormalities, Gastrointestinal Disease
ENSSSCG00000004587	MYO1E	1,47	6,80	64,45	9,90E-16	no	Metabolic Disease, Neurological Disease, Organismal Injury and Abnormalities
ENSSSCG00000025409	NMRK2	3,15	1,76	63,67	1,47E-15	no	Cellular Development, Cellular Growth and Proliferation, Cellular Movement
ENSSSCG00000006791	OVGP1	2,04	4,38	62,59	2,54E-15	no	Cellular Function and Maintenance, Cell-mediated Immune Response, Cellular Development
ENSSSCG00000009148	LEF1	-1,94	6,55	55,25	1,06E-13	no	Cellular Function and Maintenance, Cell-mediated Immune Response, Cellular Development
ENSSSCG00000004586	FAM81A	2,20	3,25	54,50	1,56E-13	no	Cellular Assembly and Organization, DNA Replication, Recombination, and Repair, Post-Translational Modification
ENSSSCG00000006758	SYT6	3,81	0,03	51,89	5,87E-13	no	Cellular Function and Maintenance, Cell-mediated Immune Response, Cellular Development



Supplementary Table S12: Genes displaying ASE in one or more tissues

MLN = Mesenteric lymph nodes / IPP = ileal Peyer's patches / JPP = Jejunal Peyer's patches / PB = Peripheral blood

(1) : List of tissues in which the gene harbouring at least one heterozygous SNP in at least one animal

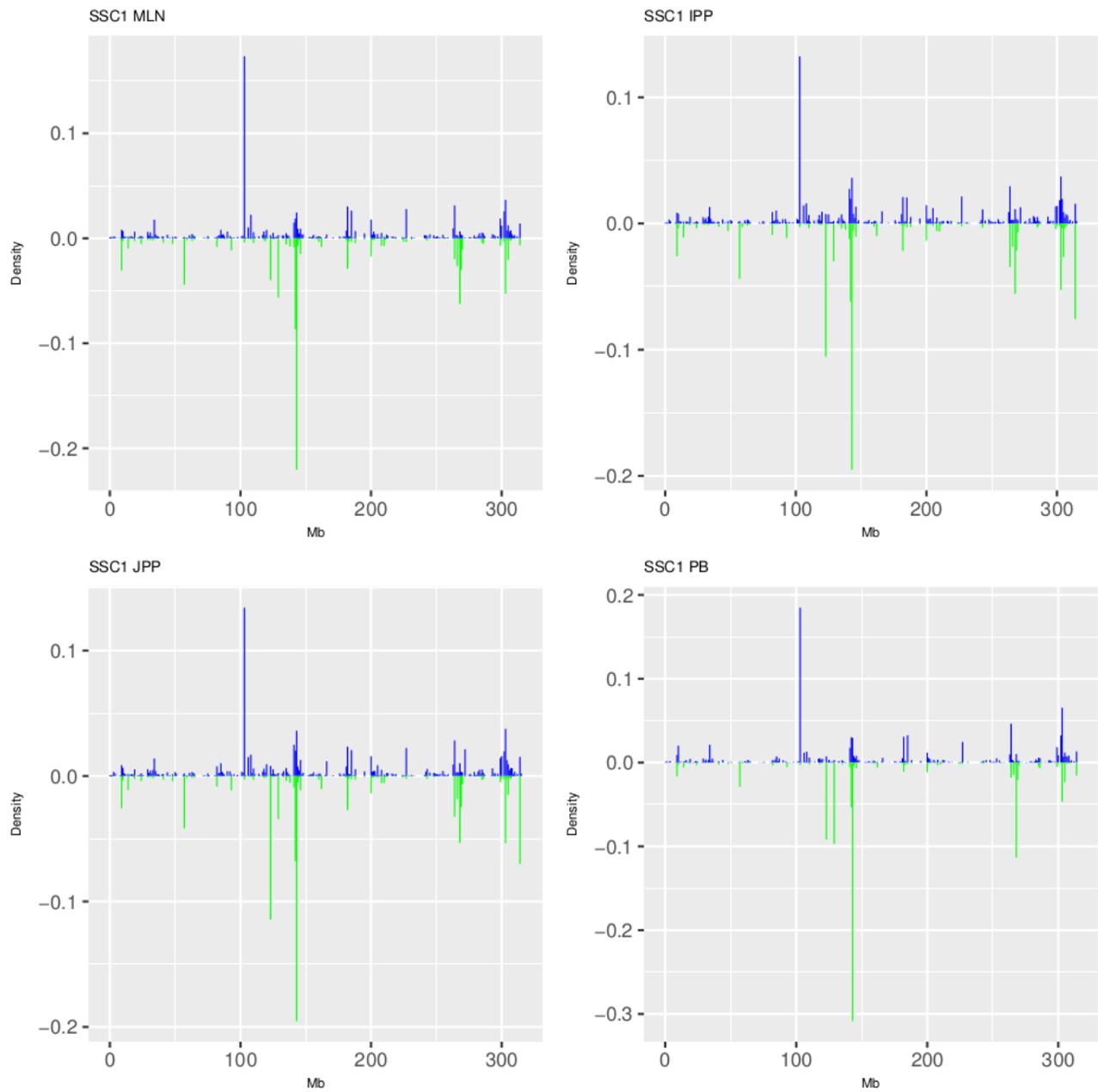
(2) : List of tissues in which the gene harbouring at least one ASE SNP in at least one animal

(3) : Number of ASE-SNPs for the gene in the tissue considered

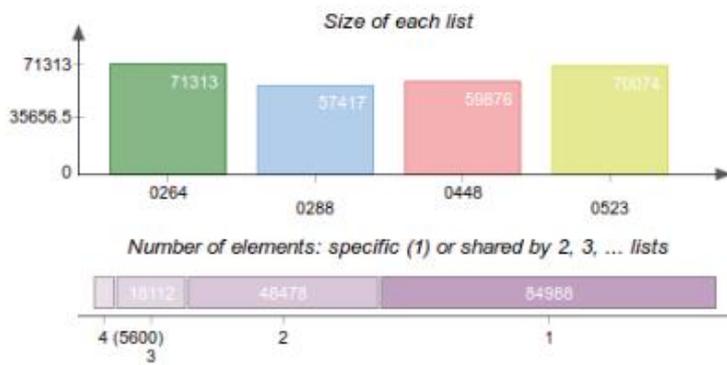
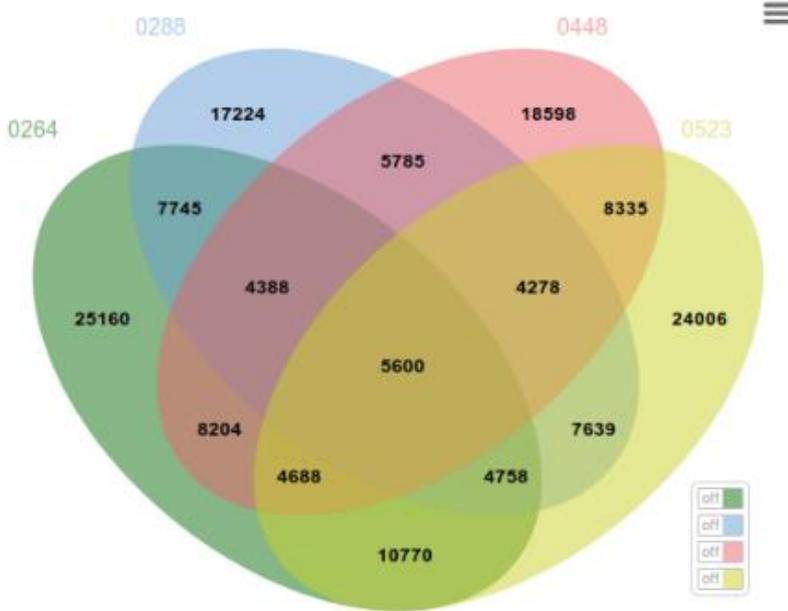
Genes with specific ASE in blood		Genes with specific ASE in IPP		Genes with specific ASE in JPP		Genes with specific ASE in MLN		
Genes with specific ASE in PPs		Genes with ASE in all tissues						
Pig_Genes_Ensembl	Pig_GeneNames	Human_GeneNames_Ortholog	HT_SNPs (1)	ASE_SNPs (2)	MLN (3)	IPP (3)	JPP (3)	PB (3)
ENSSSCG00000000007	TRMU	TRMU	MLN_IPP_JPP_PB	MLN_JPP	1	0	5	0
ENSSSCG00000000013	SMC1B	SMC1B	MLN_IPP_JPP_PB	MLN	1	0	0	0
ENSSSCG00000000018		KIAA0930	MLN_IPP_JPP_PB	MLN	2	0	0	0
ENSSSCG00000000019		NUP50	MLN_IPP_JPP_PB	IPP	0	1	0	0
ENSSSCG00000000025	PARVG	PARVG	MLN_IPP_JPP_PB	MLN_IPP_JPP_PB	4	1	3	5
ENSSSCG00000000031	MCAT	MCAT	MLN_IPP_JPP_PB	IPP	0	1	0	0
ENSSSCG00000000033	TSPO	TSPO	MLN_IPP_JPP_PB	IPP_JPP	0	3	1	0
ENSSSCG00000000034		TTL12	MLN_IPP_JPP_PB	IPP_JPP	0	1	1	0
ENSSSCG00000000036	PACSIN2	PACSIN2	MLN_IPP_JPP_PB	MLN_IPP_JPP	1	1	1	0
ENSSSCG00000000045		NDUFA6	MLN_IPP_JPP_PB	MLN_IPP_JPP_PB	1	1	1	1
ENSSSCG00000000058	SNU13	SNU13	MLN_IPP_JPP_PB	MLN_IPP_JPP_PB	1	1	1	1
ENSSSCG00000000059	XRCC6	XRCC6	MLN_IPP_JPP_PB	MLN_IPP_JPP_PB	1	1	1	1
ENSSSCG00000000060	DESI1	DESI1	MLN_IPP_JPP_PB	MLN_IPP_PB	4	1	0	3
ENSSSCG00000000075	MKL1	MKL1	MLN_IPP_JPP_PB	MLN_PB	1	0	0	1
ENSSSCG00000000076	SGSM3	SGSM3	MLN_IPP_JPP_PB	PB	0	0	0	1
ENSSSCG00000000077		AL022238.4	MLN_IPP_JPP_PB	MLN	1	0	0	0
ENSSSCG00000000079		FAM83F	IPP_JPP	IPP_JPP	0	7	8	0



Supplementary Figure S1: Density of sense and antisense reads aligned along the reference genome



Supplementary Figure S2: Comparison of heterozygous SNPs



Bibliography

1. Abreu, M. T. Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. *Nat. Rev. Immunol.* **10**, 131–144 (2010).
2. Buettner, M. & Bode, U. Lymph node dissection--understanding the immunological function of lymph nodes. *Clin. Exp. Immunol.* **169**, 205–212 (2012).
3. Reboldi, A. & Cyster, J. G. Peyer's patches: organizing B-cell responses at the intestinal frontier. *Immunol. Rev.* **271**, 230–245 (2016).
4. Chu, R. M. & Liu, C. H. Morphological and functional comparisons of Peyer's patches in different parts of the swine small intestine. *Vet. Immunol. Immunopathol.* **6**, 391–403 (1984).
5. Rothkötter, H. J., Pabst, R. & Bailey, M. Lymphocyte migration in the intestinal mucosa: entry, transit and emigration of lymphoid cells and the influence of antigen. *Vet. Immunol. Immunopathol.* **72**, 157–165 (1999).
6. Bailey, M., Christoforidou, Z. & Lewis, M. C. The evolutionary basis for differences between the immune systems of man, mouse, pig and ruminants. *Vet. Immunol. Immunopathol.* **152**, 13–19 (2013).
7. Andersen, J. K. *et al.* Systematic characterization of porcine ileal Peyer's patch, I. apoptosis-sensitive immature B cells are the predominant cell type. *Immunology* **98**, 612–621 (1999).
8. Yamamoto, M. *et al.* Alternate mucosal immune system: organized Peyer's patches are not required for IgA responses in the gastrointestinal tract. *J. Immunol.* **164**, 5184–5191 (2000).
9. Levast, B. *et al.* Differences in transcriptomic profile and IgA repertoire between jejunal and ileal Peyer's patches. *Dev. Comp. Immunol.* **34**, 102–106 (2010).
10. Gourbeyre, P. *et al.* Pattern recognition receptors in the gut: analysis of their expression along the intestinal tract and the crypt/villus axis. *Physiol Rep* **3**, (2015).
11. Mach, N. *et al.* Extensive expression differences along porcine small intestine evidenced by transcriptome sequencing. *PLoS ONE* **9**, e88515 (2014).
12. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
13. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* **44**, D710–716 (2016).
14. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
15. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
16. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25 (2010).
17. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).



18. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-33 (2013).
19. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
20. Harvey, C. T. *et al.* QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics* **31**, 1235–1242 (2015).
21. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
22. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
23. Lunney, J. K. Advances in swine biomedical model genomics. *Int. J. Biol. Sci.* **3**, 179–184 (2007).
24. Rothkötter, H.-J. Anatomical particularities of the porcine immune system--a physician's view. *Dev. Comp. Immunol.* **33**, 267–272 (2009).
25. Boeker, M., Pabst, R. & Rothkötter, H. J. Quantification of B, T and null lymphocyte subpopulations in the blood and lymphoid organs of the pig. *Immunobiology* **201**, 74–87 (1999).
26. Dvorak, C. M. T. *et al.* Gene discovery and expression profiling in porcine Peyer's patch. *Vet. Immunol. Immunopathol.* **105**, 301–315 (2005).
27. Machado, J. G., Hyland, K. A., Dvorak, C. M. T. & Murtaugh, M. P. Gene expression profiling of jejunal Peyer's patches in juvenile and adult pigs. *Mamm. Genome* **16**, 599–612 (2005).
28. Rothkötter, H. J., Huber, T., Barman, N. N. & Pabst, R. Lymphoid cells in afferent and efferent intestinal lymph: lymphocyte subpopulations and cell migration. *Clin. Exp. Immunol.* **92**, 317–322 (1993).
29. Rothkötter, H. J., Ulbrich, H. & Pabst, R. The postnatal development of gut lamina propria lymphocytes: number, proliferation, and T and B cell subsets in conventional and germ-free pigs. *Pediatr. Res.* **29**, 237–242 (1991).
30. Okada, T. *et al.* Chemokine requirements for B cell entry to lymph nodes and Peyer's patches. *J. Exp. Med.* **196**, 65–75 (2002).
31. Agace, W. W. & McCoy, K. D. Regionalized Development and Maintenance of the Intestinal Adaptive Immune Landscape. *Immunity* **46**, 532–548 (2017).
32. Jago, C. B., Yates, J., Câmara, N. O. S., Lechler, R. I. & Lombardi, G. Differential expression of CTLA-4 among T cell subsets. *Clin. Exp. Immunol.* **136**, 463–471 (2004).
33. Zabel, B. A. *et al.* Human G protein-coupled receptor GPR-9-6/CC chemokine receptor 9 is selectively expressed on intestinal homing T lymphocytes, mucosal lymphocytes, and thymocytes and is required for thymus-expressed chemokine-mediated chemotaxis. *J. Exp. Med.* **190**, 1241–1256 (1999).
34. O'Shea, J. J. & Paul, W. E. Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* **327**, 1098–1102 (2010).
35. Ma, C. S. & Phan, T. G. Here, there and everywhere: T follicular helper cells on the move. *Immunology* **152**, 382–387 (2017).



36. Ramiscal, R. R. & Vinuesa, C. G. T-cell subsets in the germinal center. *Immunol. Rev.* **252**, 146–155 (2013).
37. Tsuji, M. *et al.* Preferential generation of follicular B helper T cells from Foxp3+ T cells in gut Peyer's patches. *Science* **323**, 1488–1492 (2009).
38. Perruzza, L. *et al.* T Follicular Helper Cells Promote a Beneficial Gut Ecosystem for Host Metabolic Homeostasis by Sensing Microbiota-Derived Extracellular ATP. *Cell Rep* **18**, 2566–2575 (2017).
39. Rothkötter, H. J., Geist, M., Fritz, F. J. & Pabst, R. Age-dependence of lymphocyte production in Peyer's patch follicles in contrast to the other Peyer's patch compartments and the thymus. *Adv. Exp. Med. Biol.* **237**, 81–85 (1988).
40. Müller-Schoop, J. W. & Good, R. A. Functional studies of Peyer's patches: evidence for their participation in intestinal immune responses. *J. Immunol.* **114**, 1757–1760 (1975).
41. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzler, K. W. The antisense transcriptomes of human cells. *Science* **322**, 1855–1857 (2008).
42. Chen, C. *et al.* A genome-wide investigation of expression characteristics of natural antisense transcripts in liver and muscle samples of pigs. *PLoS ONE* **7**, e52433 (2012).
43. Chamberlain, A. J. *et al.* Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* **16**, 993 (2015).
44. Crowley, J. J. *et al.* Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* **47**, 353–360 (2015).
45. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
46. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
47. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).



III. Etude 3 : Intégration de profils transcriptomiques sanguins et phénotypes immunitaires pour l'identification de biomarqueurs sanguins et de marqueurs génétiques pour des caractères immunitaires

Article en préparation

A. Introduction

Le contrôle des maladies en élevage représente un enjeu majeur pour les systèmes de production porcines comme pour d'autres espèces. Les pathologies impactent directement le bien-être animal, la productivité des élevages, la sécurité sanitaire des produits, et la santé publique en cas de zoonoses. De plus, l'infection par un pathogène entraîne souvent un affaiblissement de l'animal qui est alors plus sensible à des infections secondaires. L'utilisation d'antibiotiques, les mesures d'hygiène mises en place et les vaccinations ne suffisent pas actuellement à contrôler pleinement le statut sanitaire des élevages de production, notamment dans un contexte où il faut réduire l'usage des antibiotiques pour limiter la diffusion des résistances (Plan ECOANTIBIO 1 et 2), et où l'efficacité vaccinale est parfois insuffisante, comme par exemple celle du vaccin contre le SDRP. Sélectionner des animaux pour leur immunocompétence est identifié comme une piste d'action, afin d'augmenter la résistance des troupeaux aux maladies infectieuses, en améliorant leur robustesse et leur résilience, c'est-à-dire leur capacité à maintenir de bons niveaux de performances même en cas d'infection.

Il existe deux axes complémentaires de recherche : (i) étudier la variabilité individuelle des réponses aux infections et repérer les animaux sensibles et résistants, (ii) étudier de manière non ciblée sur un pathogène donné la compétence des animaux et identifier des paramètres biologiques qui pourraient prédire la capacité du système immunitaire à se mobiliser lors d'une infection et donc la capacité individuelle de se protéger. Il n'est pas évident de sélectionner les animaux directement en élevage sur leur résistance aux maladies car les challenges possibles sont nombreux et varient en fonction de l'environnement. La mesure de caractères spécifiques des mécanismes de l'immunité est une approche alternative permettant d'évaluer le statut immunitaire d'un individu à un temps donné (Calder, 2007 ; Duffy *et al.*, 2014).

A tout instant et pour tout type de réaction immunitaire, le sang est un acteur majeur permettant à la réaction immunitaire de se mettre en place là où elle est nécessaire, de façon très spécifique, par le transit des différents effecteurs sur le lieu de l'inflammation. Ainsi, les paramètres immunitaires du sang constituent un groupe intéressant de phénotypes pour caractériser l'immunité, soit par mesures directes de paramètres



sanguins, soit par des tests fonctionnels *ex vivo*, comme présenté dans le Tableau Annexe 2. De vastes études sont ainsi conduites pour mesurer à grande échelle un large éventail de paramètres immunitaire à partir du sang chez l'homme (ex : projets Milieu Intérieur (Thomas *et al.*, 2015), ImmVar (De Jager *et al.*, 2015)), comme chez le porc (ex : projets IMMOPIG, SUS_FLORA). Plusieurs études ont révélé une héritabilité globalement modérée de multiples caractères immunitaires innés ou adaptatifs mesurés à partir d'échantillons sanguins : taux de sous-populations leucocytaires (Clapperton *et al.*, 2008, 2009), comme le taux de cellules tueuses (NK pour Natural Killer) et de lymphocytes T tueurs (NKT pour Natural Killer T) (Denyer *et al.*, 2006; Haverson *et al.*, 2001) ou niveaux de protéines inflammatoires de la phase aqueuse (Clapperton *et al.*, 2009). Le projet IMMOPIG a permis d'estimer l'héritabilité de nombreux paramètres immunitaires à partir d'échantillons sanguins avec ou sans stimulation chez des porcs Large-White âgés de 60 jours (Flori *et al.*, 2011). Dans cette étude, 54 paramètres immunitaires ont été étudiés sur une population de 443 porcs vaccinés contre *Mycoplasma hyopneumoniae*. La liste des paramètres étudiés incluait des anticorps spécifiques et non spécifiques, des protéines inflammatoires sériques, des sous-populations cellulaires par hémogramme et cytométrie de flux, la production *ex vivo* de cytokines (IFN- α , TNF- α , IL-6, IL-8, IL-12, IFN- γ , IL-2, IL-4, IL-10), la phagocytose et la prolifération lymphocytaire. Bien que six caractères immunitaires aient des héritabilités faibles ou non significativement différentes de zéro, les héritabilités estimées étaient modérées ($0,1 < h^2 < 0,4$) pour 18 paramètres et élevées pour 30 autres ($h^2 > 0,4$). Les corrélations phénotypiques et génétiques entre les paramètres étaient faibles, à l'exception de quelques traits qui impliquent principalement des sous-populations de cellules. IMMOPIG a permis également d'identifier des biomarqueurs sanguins candidats en comparant le transcriptome d'individus à valeurs phénotypiques extrêmes (Mach *et al.*, 2013).

La présente étude s'appuie sur les données du projet SUS_FLORA avec l'analyse de la variation de 55 paramètres immunitaires sanguins dans un nouveau groupe de 550 porcs Large-White, phénotypés à 60 jours. Cette nouvelle étude nous permet de consolider et enrichir les résultats obtenus avec les données du projet IMMOPIG, en estimant les paramètres génétiques de 55 caractères immunitaires sanguins, dont 30 représentent des phénotypes descriptifs (numération de formule sanguine, identification de sous-populations leucocytaires, dosages sériques d'anticorps et de protéines inflammatoires) et 25 autres des phénotypes de réponse à des stimulations *ex vivo* (capacité de phagocytose, production de cytokines après stimulation inflammatoire).

Pour un sous-groupe de 243 individus pour lesquels le transcriptome sanguin a été analysé par hybridation de puces d'expression, nous avons identifié des gènes dont la variation d'expression dans le sang



est associée à la variation d'au moins un caractère immunitaire. Ces gènes constituent un ensemble de biomarqueurs candidats pour chaque caractère pris indépendamment. Des approches multivariées nous ont permis de prioriser des gènes qui permettraient de prédire la variation de plusieurs caractères en même temps. Enfin, nous avons regardé quels biomarqueurs candidats étaient décrits comme étant sous contrôle génétique, en comparant nos résultats avec ceux obtenus dans le Chapitre 1 de cette thèse.

B. Matériels et Méthodes

1. Animaux et échantillonnage

Les animaux utilisés dans cette étude proviennent d'une cohorte de 550 porcs Large White produits durant le projet SUS_FLORA. Du sang périphérique a été prélevé au niveau de la veine jugulaire chez les porcs âgés de 60 jours et collecté sur différents types de tubes selon les applications : tube PAXgene® (extraction d'ARN), tube sec (dosages sériques), tube EDTA (formule sanguine, extraction d'ADN), et tube héparinisé (tests *ex vivo*, étude des sous-populations cellulaires par cytométrie en flux). Le sang collecté sur tube sec a été centrifugé à 3200g pendant 15min à 4°C. Le sérum a ensuite été récupéré et conservé à -20°C jusqu'à utilisation. Le sang collecté sur les tubes PAXgene® a été conservé à -80°C. Après avoir réalisé les numérations-formule sanguine, le sang collecté dans les tubes EDTA a été conservé à -20°C. Le sang collecté dans des tubes héparinisés a été stocké à température ambiante et exploité le lendemain pour les expériences de phagocytose et de stimulation de sang total. Le protocole d'expérimentation animal a été approuvé par le comité d'éthique COMETHEA Poitou-Charentes avec l'autorisation CE2013-2.

2. Extraction d'ARN et analyse du transcriptome

Les données d'expression des 243 animaux utilisées dans ce travail, ont été déjà décrites par Maroilley *et al.* (article en révision ; Chapitre 1 de cette thèse). L'ARN du sang a été extrait à partir des tubes PAXgene® en utilisant le kit PAXgene® Blood RNA kit (QIAGEN), selon les instructions du fabricant. La concentration, la pureté et la qualité des ARN ont été évaluées par spectrophotométrie avec un Nanodrop 2000 (ThermoFisher Scientific) et par électrophorèse sur un Bioanalyseur 2100 (Agilent) en utilisant le kit Eukaryote total RNA 6000 Nano. La moyenne du rapport 260/280, déterminant la pureté de l'ARN a été de 2,14 ($\pm 0,03$) et celle du RIN (pour *RNA Integrity Number*), mesurant de qualité de l'ARN, de 7,31 (+/-0,7).



Le protocole d'analyse du transcriptome sanguin par hybridation de microarrays a déjà été décrit dans Maroilley *et al.* (en révision, voir Chapitre 1, numéro d'accèsion : GSE97374). En bref, la normalisation, la filtration et le contrôle qualité ont été réalisés avec le package R Limma sur les 60 306 sondes. Les valeurs d'expression ont été normalisées par la méthode des quantiles. Les sondes ont été filtrées selon leur intensité afin d'éliminer les gènes sous-exprimés : seules les sondes s'exprimant à des niveaux au moins 10% supérieurs à ceux des contrôles négatifs pour au moins quatre échantillons ont été retenues. Au total, 59 774 sondes ayant passé les contrôles de qualité et les étapes de filtration ont été considérées comme exprimées dans les échantillons sanguins inclus dans cette étude. La variation d'expression des sondes a été corrigée par un modèle linéaire incluant les effets du sexe et de la bande des animaux.

L'annotation des séquences des sondes a été précisée en s'appuyant sur le génome de référence *Sus scrofa* version 11.1 et l'annotation disponible dans la version 90 de la base de données Ensembl, en utilisant la même approche décrite pour la version 10.2 du génome porcin (Chapitre 1, Maroilley *et al.*, en révision).

3. Phénotypage des traits immunitaires

a) Numérations-formules sanguines (NFS)

Les NFS ont été réalisées avec un hémogramme MS4-5 (ElitechGroup, France) à partir de sang prélevé sur tube EDTA. Les 12 paramètres mesurés sont reportés dans le Tableau à l'Annexe 2. Nous avons également calculé le ratio granulocytes / lymphocytes (GRA_LYM), souvent utilisé comme un indicateur de santé et de stress (Al-Murrani *et al.*, 2006 ; Davis *et al.*, 2008).

b) Niveaux sériques d'haptoglobine, de protéine C-réactive, d'IgA totales et d'IgG spécifiques de *Mycoplasma hyopneumoniae*

L'haptoglobine, la protéine C-réactive (C-Reactive Protein, CRP) et les taux d'IgA ont été mesurés dans le sérum des porcs en utilisant des kits commerciaux (Phase Haptoglobin Assay de chez ABCYS Biologie, Pig C-Reactive Protein ELISA Kit de chez Genway Biotech Inc et Pig IgA ELISA Quantification Kit de chez Bethyl Laboratories, respectivement). Les dosages ont été réalisés en duplicat selon les instructions des fournisseurs. Les échantillons n'ont pas été dilués pour le dosage de l'haptoglobine, et dilués au 1/2000 pour les dosages de CRP et d'IgA. Les résultats sont exprimés en mg/mL.

Les taux sériques d'IgG spécifiques de *Mycoplasma hyopneumoniae* (IgG_MH) ont été mesurés avec un kit ELISA commercial (ID Screen® *Mycoplasma hyopneumoniae* Indirect de chez IDVET) en diluant les échantillons au 1/40 au lieu du 1/20 recommandé par le fabricant, car la réponse observée était très forte.



Les contrôles négatifs et positifs fournis dans le kit ont été dilués au 1/20. Les résultats sont exprimés en valeurs relatives par rapport au contrôle positif correspondant au ratio des absorbances nettes (pour lesquelles l'absorbance du contrôle négatif a été soustraite) de l'échantillon (E) sur le contrôle positif (P) (valeurs E/P).

c) Mesures de la capacité de phagocytose des leucocytes

La mesure quantitative de la phagocytose des leucocytes du sang héparinisé a été réalisée en utilisant le kit PHAGOTEST (Glycotope) selon les recommandations du fournisseur. Ce test permet de mesurer le pourcentage de monocytes et de granulocytes qui montrent une activité de phagocytose (ingestion d'une ou plusieurs bactéries par cellule) et l'activité individuelle des cellules à phagocyter (nombre de bactéries ingérées par cellule). L'acquisition des données a été réalisée par cytométrie en flux (FACScan de chez Becton Dickinson) et l'analyse de données a été faite grâce à un script utilisant le logiciel R (version 3.4.0) et la librairie flowCore (version 1.42.2) pour l'exécution automatique des régions et les librairies flowDensity (version 1.10.0), et flowViz (version 1.40.0) pour la visualisation des données. La stratégie d'analyse est présentée sur la Figure E3-28. Les résultats sont exprimés en % de cellules phagocytaires rapportés aux cellules totales du sang, aux granulocytes ou aux monocytes, et en mesure de l'intensité moyenne de fluorescence (MFI) des cellules phagocytaires, ce niveau correspondant à la quantité de bactéries internalisées (phagocytées).



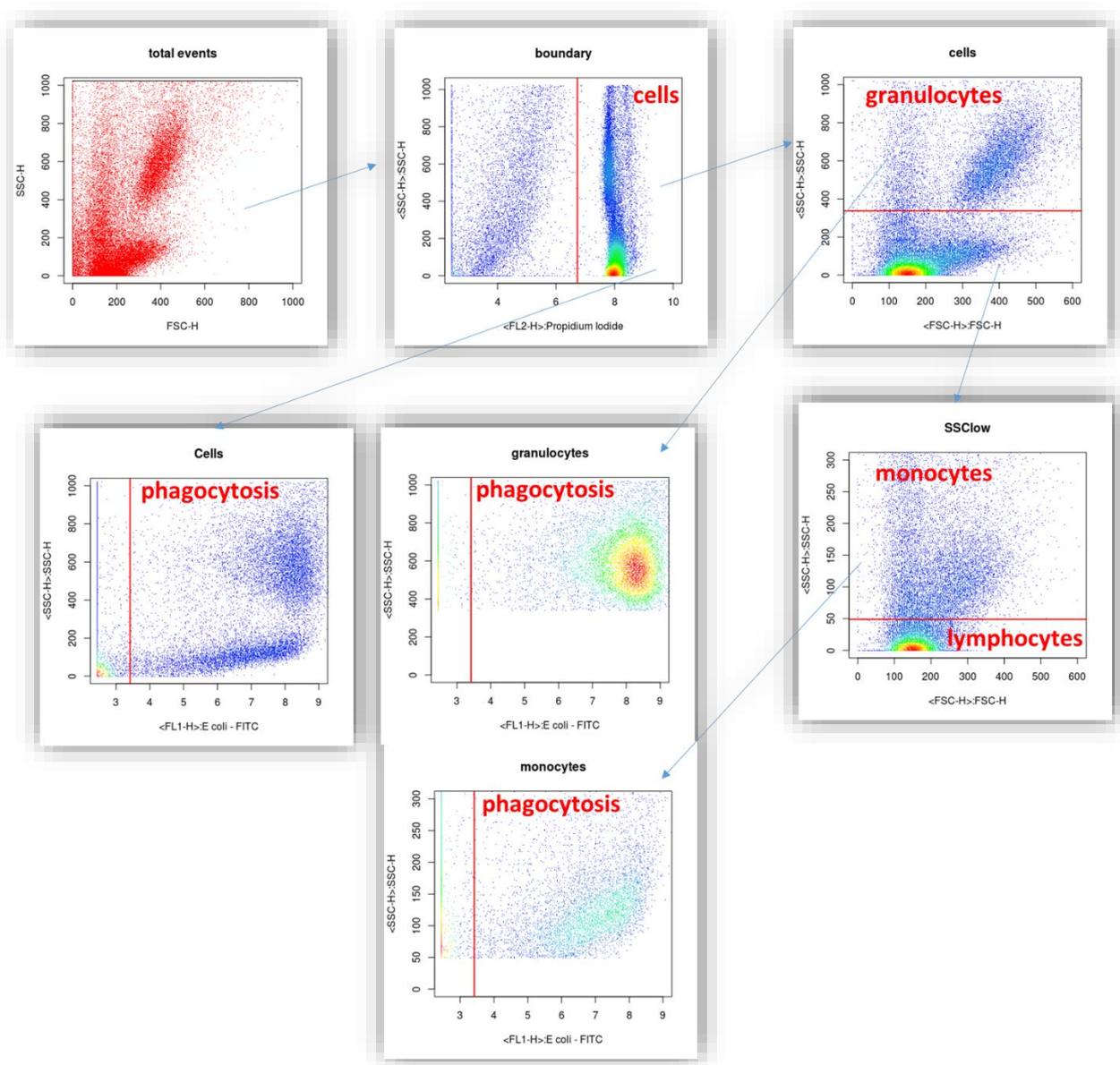


Figure E3-28 : Stratégie d'analyse des données de cytométrie en flux pour analyser la capacité de phagocytose. Les événements en limite de détection du cytomètre (boundary events) sont exclus, dans le premier dot plot FSC (Forward scatter, taille des cellules) vs SSC (side scatter, granulométrie des cellules), seuls ceux en rouge sont conservés. Dans le deuxième dot plot représentant le marquage par l'iodure de propidium vs FSC, les cellules sanguines sont identifiées par leur contenu en ADN caractéristique d'une cellule diploïde (« cells »). Parmi ces cellules, dans le troisième dot plot, le profil SSC vs SSC permet d'identifier les cellules fortement granuleuses (« granulocytes »), les autres (« SSC low ») sont ensuite à nouveau séparées selon leur taille et granulométrie, les moins granuleuses correspondant aux « lymphocytes » et les autres aux « monocytes ». La phagocytose est ensuite évaluée pour les cellules totales (« cells ») ou au sein des « granulocytes » ou des « monocytes » en étudiant la fluorescence en FITC, les bactéries *E coli* utilisées dans ce test étant marquées par le FITC. Les cellules FITC+ sont celles qui ont phagocyté les bactéries *E coli*.

d) Quantification des sous-populations cellulaires dans les PBMC par cytométrie en flux

Les PBMC (Peripheral Blood Mononuclear Cells ou cellules mononuclées du sang périphérique) ont été purifiées par centrifugation sur gradient de densité. Ainsi, 13mL de sang héparinisé ont été déposés dans un tube Leucosept (Greiner Bio-one) pré-rempli avec 17mL de Ficoll (Lymphocytes Separation Medium, Eurobio). Après centrifugation à 300g pendant 35min, l'anneau de PBMC situé à l'interface a été prélevé et lavé en D-PBS sans calcium ni magnésium (Gibco). Les érythrocytes ont ensuite été lysés en incubant les cellules dans 2mL de BD Pharmlyse (BD Biosciences) pendant 10min à température ambiante. Les PBMC ont ensuite été lavées dans 50mL de D-PBS, incubées avec 2mL de sérum porcin à 4°C pendant 20min puis lavées à nouveau dans 50mL de D-PBS puis 50mL de tampon de marquage (PBS contenant 0,1% de NaN_3 et 1% de BSA, pH 7,3). Les cellules ont finalement été reprises à une concentration finale de $5 \cdot 10^6$ cellules/mL. 10^6 cellules ont été utilisées pour chacune des trois combinaisons de marquages effectuées.

Les multiples marquages ont été faits en utilisant des anticorps monoclonaux murins. Les PBMC ont été marquées par les anticorps primaires dilués dans le tampon de marquage pendant 25min à 4°C selon les trois combinaisons de marquages : (i) un anti-CD16 (clone G7, isotype IgG1, 1/100, Abd Serotec) et un anti-CD8 α (clone PT81B, isotype IgG2b, 1/50, VRMD) ; (ii) un anti-TCR $\gamma\delta$ (clone PGBL22A, isotype IgG1, 1/250, VRMD) et un anti-CD8 α (clone PT81B, isotype IgG2b, 1/50, VRMD) ; (iii) un anti-CD172a (clone 74-22-15A, isotype IgG1, 1/500, VRMD) et un anti-CMH II (clone Th21a, isotype IgG2b, 1/250, VRMD).

Les cellules ont ensuite été lavées dans le tampon de marquage puis incubées pendant 25min à 4°C avec les anticorps secondaires dilués au 1/200 dans le tampon de marquage : un anti-IgG1 de souris couplé au Tricolor (Invitrogen) et un anti-IgG2b de souris couplé à la phycoérythrine (Invitrogen). Les cellules ont à nouveau été lavées dans le tampon de marquage puis incubées pendant 25min à 4°C avec les anticorps primaires directement conjugués à des fluorochromes dilués dans le tampon de marquage selon les trois combinaisons de marquages : (i) un anti-CD3-FITC (clone PPT3, 1/250, Abcam) ; (ii) un anti-CD3-FITC (clone PPT3, 1/250, Abcam) et un anti-CD4-PE (clone 74-12-4, 1/250, Abcam) ; (iii) un anti-CD14-FITC (clone Tük4, 1/50, Abcam) et un anti-CD163-PE (clone 2A/10/11, 1/20, Abd Serotec). Après un dernier lavage dans le tampon de marquage, les cellules ont finalement été fixées dans la solution de fixation BD Cellfix (BD Biosciences).

L'acquisition des données a été réalisée avec un cytomètre en flux (FACScan de chez Becton Dickinson) et l'analyse de données a été faite grâce à un script utilisant le logiciel R (version 3.4.0) et les bibliothèques flowCore (version 1.42.2) et openCyto (version 1.14.0) pour l'exécution automatique des régions et les bibliothèques flowDensity (version 1.10.0), et flowViz (version 1.40.0) pour la visualisation des données. Les



stratégies d'analyse des données de cytométrie en flux pour les trois combinaisons de marquage sont illustrées dans les Figures E3-29, E3-30 et E3-31 pour les combinaisons de marquages i, ii et iii, respectivement.

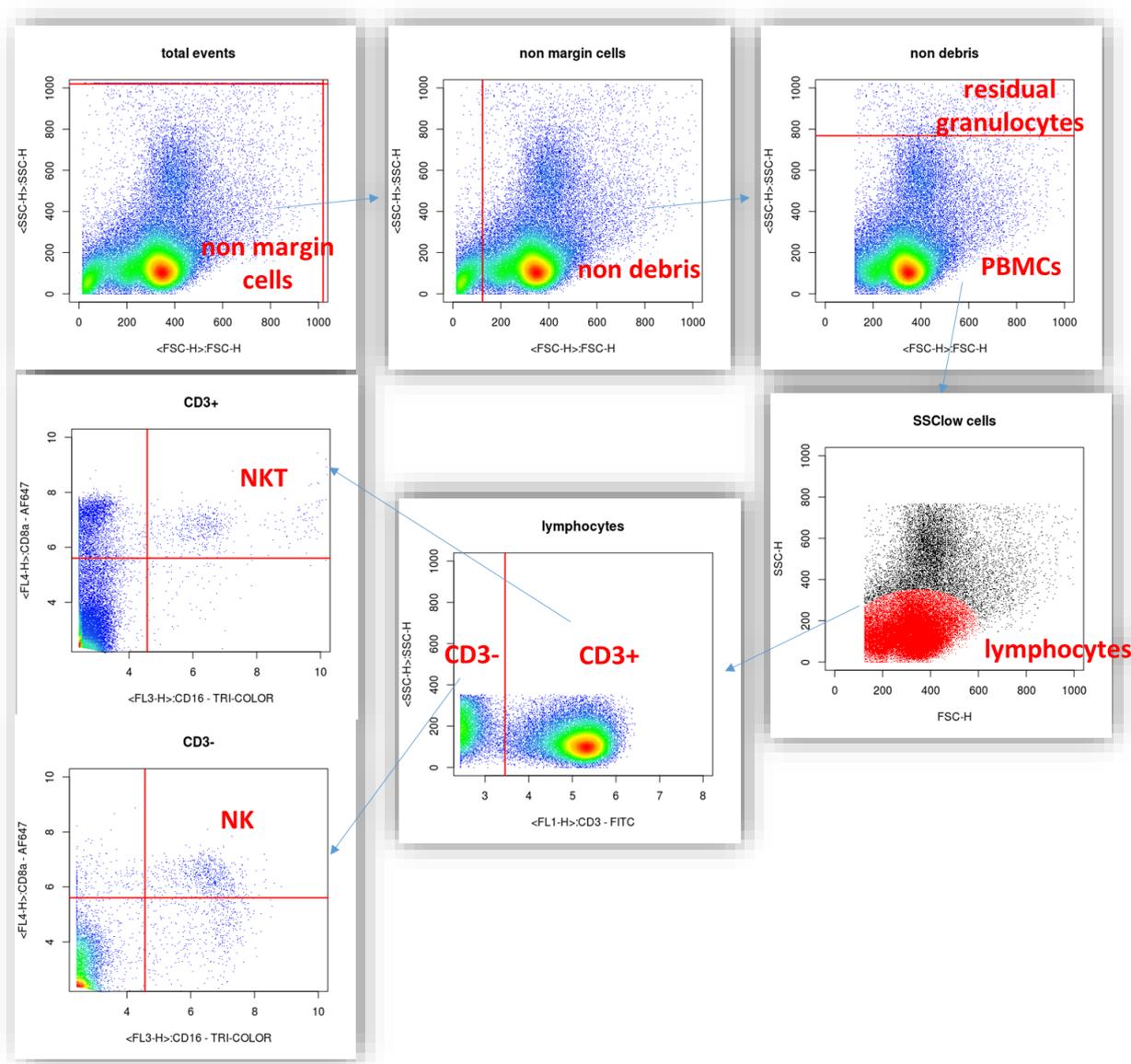


Figure E3-29 : Stratégie d'analyse des données de cytométrie en flux pour identifier les cellules NK et NKT dans les PBMC du sang de porc. Sur le premier dot plot FSC vs SSC, les événements en limite de détection du cytomètre sont exclus, seuls les « non margin cells » sont conservées. Dans le deuxième dot plot, les débris (FSC très faible) sont exclus et les « non debris » sont ensuite représentés en FSC vs SSC, les granulocytes résiduels (ayant une granulométrie importante) sont également exclus. Parmi les « PBMCs », deux populations sont clusterisées selon leur profil FSC vs SSC, les cellules les plus petites et non granuleuses sont identifiées comme étant les « lymphocytes » (en rouge sur le dot plot). Parmi ces « lymphocytes », le marquage CD3 est analysé et les cellules « CD3- » et « CD3+ » sont identifiées. Sur chacune de ces populations, un dot plot CD16 vs CD8 α est tracé. Les cellules CD8 α + CD16+ sont identifiées comme les cellules « NK » à partir des cellules « CD3- » et les cellules « NKT » à partir des cellules « CD3+ ».



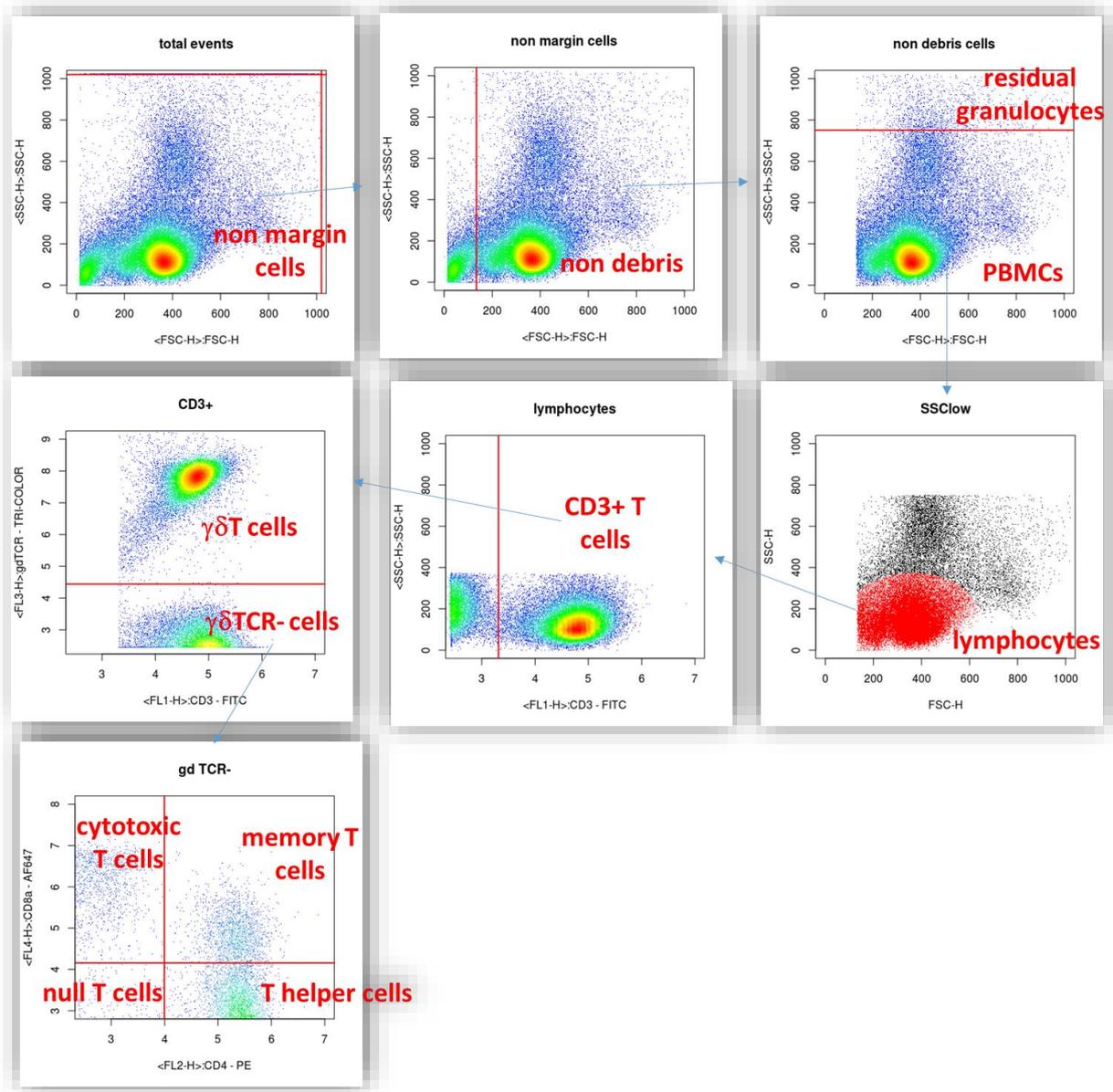


Figure E3-30 : Stratégie d'analyse des données de cytométrie en flux pour identifier les lymphocytes T $\gamma\delta$, les lymphocytes T cytotoxiques (T CD4-CD8+), les lymphocytes T helper CD4 (CD4+ CD8-) et les lymphocytes T mémoire (T CD4+CD8+) dans les PBMC du sang de porc. Sur le premier dot plot FSC vs SSC, les événements en limite de détection du cytomètre sont exclus, seuls les « non margin cells » sont conservées. Dans le deuxième dot plot, les débris (FSC très faible) sont exclus et les « non debris » sont ensuite représentés en FSC vs SSC, les granulocytes résiduels (ayant une granulométrie importante) sont également exclus. Parmi les « PBMCs », deux populations sont clusterisées selon leur profil FSC vs SSC, les cellules les plus petites et non granuleuses sont identifiées comme étant les « lymphocytes » (en rouge sur le dot plot). Parmi ces « lymphocytes », le marquage CD3 est analysé et les lymphocytes T sont alors identifiés (« CD3+ T cells »). Un dot plot CD3 vs TCR $\gamma\delta$ est tracé, les cellules exprimant le TCR $\gamma\delta$ sont les lymphocytes T $\gamma\delta$ (« $\gamma\delta$ T cells », celles ne l'exprimant pas (« $\gamma\delta$ TCR- cells ») sont ensuite représentées dans le dot plot CD4 vs CD8 α . Quatre sous-populations sont alors identifiées : les cellules CD4- CD8 α + correspondent aux lymphocytes T cytotoxiques (« cytotoxic T cells »), les CD4+ CD8 α - aux lymphocytes T helper (« T helper cells »), les CD4+ CD8 α + aux lymphocytes T mémoire (« memory T cells ») et les CD4- CD8 α - à des lymphocytes T immatures (« null T cells »).

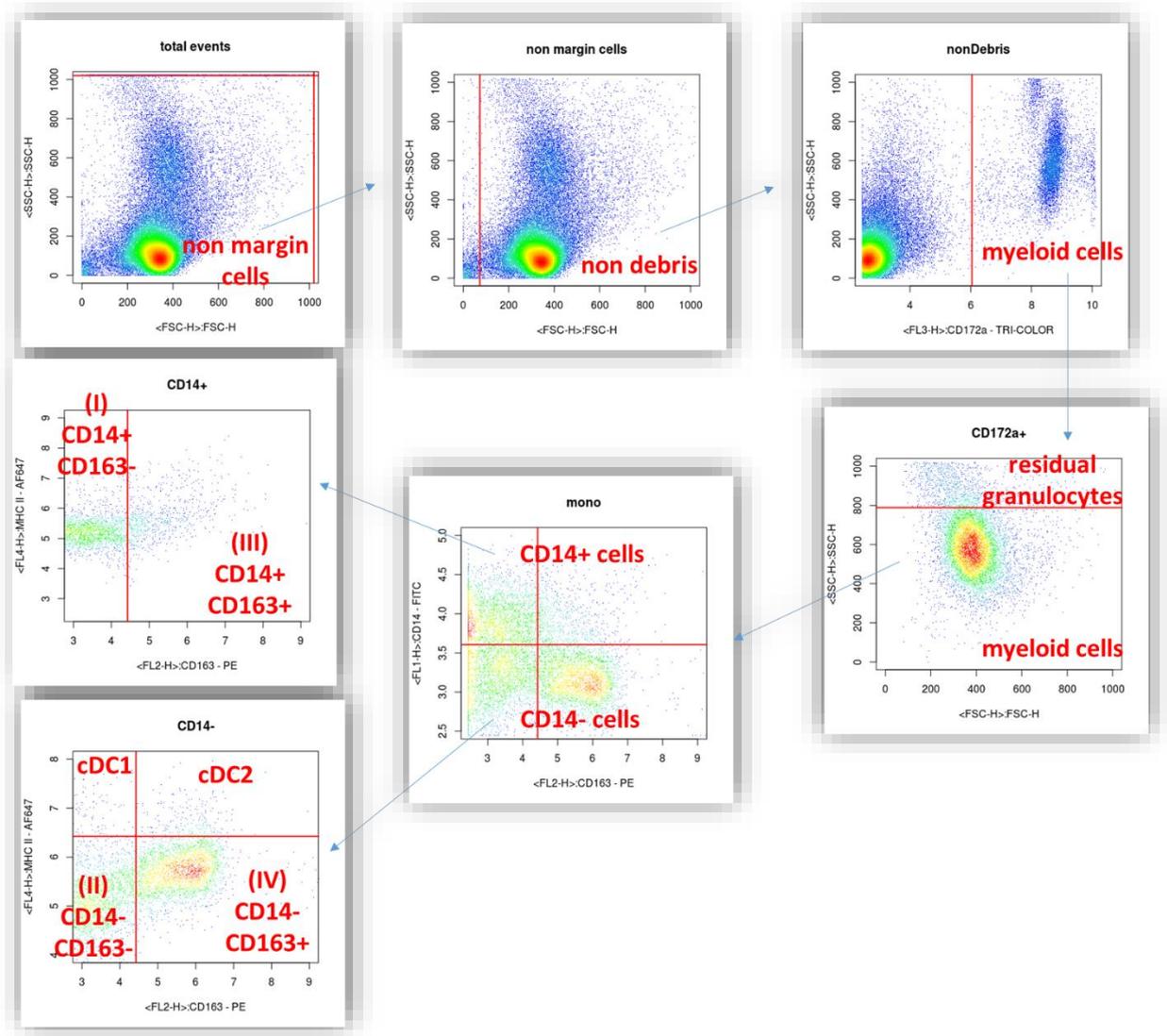


Figure E3-31 : Stratégie d'analyse des données de cytométrie en flux pour identifier les cellules dendritiques (cDC1 et cDC2), les 4 types de monocytes (I à IV désignant des stades de maturation croissants) dans les PBMC du sang de porc. Sur le premier dot plot FSC vs SSC, les événements en limite de détection du cytomètre sont exclus, seuls les « non margin cells » sont conservées. Dans le deuxième dot plot, les débris (FSC très faible) sont exclus et les « non debris » sont ensuite représentés sur le dot plot CD172a vs SSC. Les cellules myéloïdes, exprimant CD172a, sont alors sélectionnées (« myeloid cells »). Les granulocytes résiduels parmi ces cellules sont exclus dans le dot plot FSC vs SSC. Elles correspondent aux cellules ayant une forte granulométries (« residual granulocytes »). Le marquage CD14 est ensuite analysé parmi les cellules myéloïdes restantes, les cellules « CD14+ » et « CD14- » sont ensuite analysées. Les dot plot CD163 vs CMH II sont tracés pour ces deux sous-populations. Les cellules « CD14+ CD163- » et « CD14+ CD163+ » sont identifiées comme étant deux sous-populations de monocytes. Parmi les cellules « CD14- », les cellules dendritiques sont identifiées comme exprimant très fortement les CMH II, les « cDC1 » étant CD163- et les « cDC2 » CD163+. Les autres sous-populations identifiées sont alors deux autres sous-populations de monocytes : « CD14- CD163- » et « CD14- CD163+ ».

e) Stimulation *in vitro* des cellules du sang

Une quantité de 400µL de sang collecté sur tube héparinisé a été déposée dans les puits d'une plaque 24 puits avec 1,6mL de milieu RPMI 1640 (Bio Whittaker) contenant 10% de sérum de veau fœtal décomplémenté (QB Perbio), 2 mM de L-glutamine, 100 U/mL de pénicilline et 100 mg/mL de streptomycine. Pour la stimulation, un mélange de 10 ng/ml de PMA (Sigma), 1µg/mL de ionomycine (Sigma) et/ou 1 µg/mL de LPS provenant d'*Escherichia coli* O111:B4 (Sigma) a été ajouté au sang dilué au 1:5. Pour les contrôles non stimulés, un volume de PBS équivalent à celui des réactifs de stimulation a été ajouté. Après incubation dans une étuve à 37°C avec une atmosphère à 5% de CO₂ pendant 4 ou 48h, les plaques ont été centrifugées à 450g pendant 20min et les surnageants de culture ont été prélevés et conservés à -20°C.

f) Dosage de cytokines après stimulation *in vitro* de sang total

Après 4h de stimulation PIL (PMA + Ionomycine + LPS), les concentrations d'IL-1β, d'IL-2, d'IL-8, d'IL-10, d'IL-17, de TNF-α et d'INF-γ ont été dosées en multiplex avec des billes (Cytokine Beads Assay, CBA) par une méthode développée au laboratoire. Les anticorps et les cytokines recombinantes utilisées sont listées dans le Tableau E3-7.

Tableau E3-7 : Liste des anticorps et cytokines recombinantes utilisés dans le test de dosage des cytokines porcines en multiplex. Les limites de quantification constatées avec ce test sont indiquées, sans tenir compte du facteur de dilution des échantillons.

<i>Cytokines</i>	<i>Anticorps de capture</i>	<i>Cytokines recombinantes</i>	<i>Anticorps biotinylés</i>	<i>Limites de quantification (pg/ml)</i>
<i>IL-1β</i>	R&D Systems MAB6811	R&D Systems 681-PI	R&D Systems BAF681	30
<i>IL-2</i>	Invitrogen ASC0924	Invitrogen PSC0024	Invitrogen ASC0829	100
<i>CXCL8/IL-8</i>	R&D Systems MAB5351	R&D Systems 535-IN	R&D Systems BAF535	14
<i>IL-10</i>	Invitrogen ASC0104	Invitrogen PSC0104	Invitrogen ASC9109	30
<i>IL-12/IL-23 p40</i>	Abd Serotec MCA2414Z	R&D Systems 912-PI	R&D Systems BAM9122	20
<i>IL-17</i>	KingfisherBiotech KP0498S	KingfisherBiotech RP0128S	KingfisherBiotech KPB0499S	4.6
<i>TNFα</i>	R&D Systems MAB6902	R&D Systems 690-PI	R&D Systems BAM6903	40
<i>IFNγ</i>	Invitrogen ASC4934	Invitrogen PSC4034	Invitrogen ASC4839	40



Les anticorps de capture ont été couplés de façon covalente à la surface de billes fluorescentes (Cytoplex™ carboxylated Microspheres de chez Thermo Fisher Scientific) selon les recommandations du fournisseur. Chaque anticorps de capture a été couplé à une microsphère différente. Tous les dosages ont été réalisés en duplicat. 50µL d'échantillons (dilués au 1/3) ou de la gamme de standards, dilués dans du tampon de dosage (137mM de NaCl, 1,35mM de KCl, 4mM de NaH₂PO₄, 1mM de KH₂PO₄, 50mM de Tris et 0,25% de BSA) contenant 0,5% de polyvinylalcool (Sigma-Aldrich) et 0,8% de polyvinylpyrrolidone (Sigma-Aldrich), ont été incubés sur une microplaque à filtre (MultiScreen_{HTS}-BV 1,2 µm, Millipore) avec 2000 billes par puits de chacune des billes couplées aux anticorps. Un lavage en trois étapes est ensuite réalisé en lavant les billes deux fois avec du tampon de lavage 10X (685mM de NaCl, 13,5mM de KCl, 40mM de NaH₂PO₄, 10mM de KH₂PO₄, 250mM de Tris) et une fois avec du tampon de lavage 1X. 25µL d'anticorps de détection biotinylés (chacun dilué à 1 µg/mL dans du tampon de dosage) sont ensuite ajoutés et incubés pendant 90min. Après lavage en trois étapes, 50µL de streptavidine couplée à la phycoérythrine (5µg/mL diluée dans du tampon de dosage, Invitrogen) ont été ajoutés dans chaque puits et incubés pendant 30min. Les billes ont finalement été lavées à nouveau puis reprises dans du tampon de dosage avant d'être acquises sur le cytomètre Guava easyCyte 6HT-2L (Millipore). Les résultats ont été analysés avec le logiciel FloCytomix Pro software. La courbe standard a été calculée en effectuant un modèle de régression logistique à cinq paramètres. Les résultats sont exprimés en pg/mL.

Après 48h de stimulation (LPS, PMA+Ionomycine (PI) et PIL), l'IL-1β, l'IL-6, l'IL-8 et le TNF-α ont été dosés dans les surnageants de culture en utilisant des tests ELISA commerciaux (DuoSet ELISA kits de chez R&D Systems) selon les instructions données par le fournisseur. Pour quantifier les niveaux basaux de cytokines dans les surnageants des cellules non stimulées, les échantillons n'ont pas été dilués. Pour les surnageants des cellules stimulées, les échantillons ont été dilués au ½ pour l'IL-1β, au ¼ pour l'IL-6, du ¼ au 1/20 pour le TNF-α et du 1/10 au 1/40 pour l'IL-8. Tous les échantillons ont été dosés en duplicate. Les limites de détection des dosages ELISA données par le fournisseur sont les suivants : 62,5 pg/mL pour l'IL-1β, 125 pg/mL pour l'IL-6, 31,2 pg/mL pour le TNF-α et 125 pg/mL pour l'IL-8, sans tenir compte du facteur de dilution des échantillons.



4. Analyses génétiques et identification de biomarqueurs

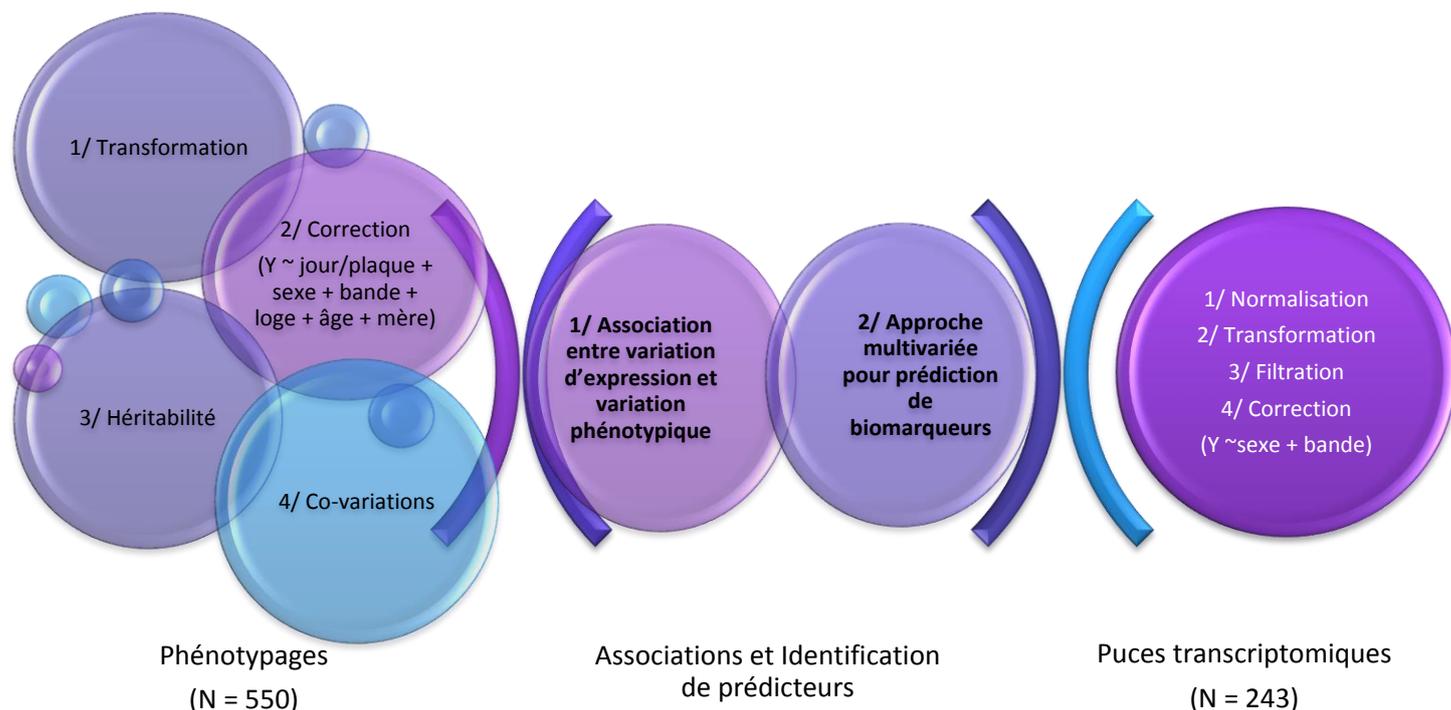


Figure E3-32 : Schéma descriptif des analyses statistiques effectuées

a) Estimation des héritabilités

L'estimation de l'héritabilité a pu être faite pour 18 caractères. Dans un premier temps, une analyse de la distribution des différents caractères a été réalisée. Une transformation logarithmique $[\log(x+1)]$ a été utilisée pour normaliser les distributions de l'ensemble des variables à l'exception de TNFa_LPS_48h. Dans certains des cas, les valeurs faibles ont été considérées comme « outliers » et éliminées des analyses finales. Une analyse de sensibilité a toutefois été réalisée de façon systématique pour apprécier l'influence de ces « outliers » sur les résultats. Le nombre de données éliminées comme « outliers » est indiqué dans le tableau ci-dessous (Tableau E3-8).

Des analyses préliminaires visant à déterminer les effets à prendre en compte dans le modèle d'estimation des paramètres génétiques ont ensuite été réalisées à l'aide de la procédure GLM du logiciel SAS. Le modèle retenu pour l'estimation des paramètres génétiques est un modèle animal prenant en compte les effets fixés de la bande de contemporains, du sexe, de la plaque utilisée (dosages des cytokines et protéines sériques), et les effets aléatoires de la portée de naissance (c), de la valeur génétique additive de chaque animal (a) et résiduels (e), ainsi que l'âge en covariable linéaire. Les variances sont supposées homogènes de sorte que les vecteurs d'effets aléatoires c et e sont proportionnels à la matrice identité et le



vecteur des effets aléatoires a est proportionnel à la matrice de parenté entre individus considérés dans l'analyse. Les covariances entre a et c , entre a et e et entre c et e sont supposées nulles. Les analyses ont été réalisées à l'aide de la version 6 du logiciel VCE (Groeneveld *et al.*, 2010; Neumaier and Groeneveld, 1998).

Tableau E3-8 : Résumé des facteurs pris en compte dans le modèle pour estimer l'héritabilité par caractère immunitaire

Caractère	Transformation	Outliers	Bande	Sexe	Manip CBA	Plaque	Age	Portée	Animal
TNFa_4h	log(x+1)		X	X	X		X	X	X
IL2_4h	log(x+1)		X	X	X		X	X	X
IL10_4h	log(x+1)	10	X	X	X		X	X	X
IL1b_4h	log(x+1)		X	X	X		X	X	X
IL8_4h	log(x+1)	4	X	X	X		X	X	X
IL17_4h	log(x+1)	3	X	X	X		X	X	X
CRP	log(x+1)		X	X		X	X	X	X
Hapto	log(x+1)		X	X		X	X	X	X
IgA	log(x+1)		X	X		X	X	X	X
IL8_LPS_48h	log(x+1)	7	X	X		X	X	X	X
TNFa_LPS_48h			X	X		X	X	X	X
IL10_PI_48h	log(x+1)	5	X	X		X	X	X	X
IL10_PIL_48h	log(x+1)		X	X		X	X	X	X
TNFa_PI_48h	log(x+1)		X	X		X	X	X	X
TNFa_PIL_48h	log(x+1)	1	X	X		X	X	X	X
IL8_PI_48h	log(x+1)	1	X	X		X	X	X	X
IL8_PIL_48h	log(x+1)	1	X	X			X	X	X

b) Corrélations entre phénotypes immunitaires et exploration des données

Pour étudier la covariation des phénotypes immunitaires sanguins, toutes les données ont été transformées par l'approche BoxCox. Les valeurs normalisées ont été corrigées par un modèle linéaire mixte, implémenté sous R par la fonction `lmer()` (Bates *et al.*, 2015), en considérant la bande, le sexe et la loge comme des effets fixes, l'âge à la vaccination comme une co-variable et la mère comme un effet aléatoire. Les résidus du modèle pour chaque phénotype ont ensuite été utilisés pour calculer la corrélation par paires de caractères immunitaires par la méthode de Spearman, en ne prenant en compte que des individus ayant des valeurs pour les deux caractères considérés.



La matrice des distances euclidiennes a été déduite de la matrice des corrélations de Spearman précédemment établie. Elle a permis une classification hiérarchique non supervisée avec la fonction `hclust()` sur R selon la méthode « complete linkage », c'est-à-dire le classement des caractères immunitaires en fonction de la similarité et de la dissimilarité de leurs valeurs. Plus les branches reliant deux caractères sont petites, plus leurs valeurs auront la même variabilité en fonction des individus.

c) Analyse d'inférence de réseaux

A partir de la matrice des corrélations de Spearman, nous avons appliqué l'algorithme de corrélation partielle, implémenté sous R dans le package PCIT (Watson-Haigh *et al.*, 2010), afin de pouvoir construire un réseau à partir des corrélations entre les caractères immunitaires, en limitant la redondance. Cet algorithme permet de déterminer si la relation de corrélation entre X et Y est directe ou partielle, expliquée par une corrélation directe entre X-Z et Y-Z. Les résultats de l'analyse du réseau obtenu ont été visualisés avec le logiciel Cytoscape.

d) Associations entre variation de transcrits dans le sang et variation de traits immunitaires

L'association entre chaque caractère immunitaire et chaque sonde d'expression a été estimée par régression multivariée par spline adaptative. Il s'agit d'une extension de la régression par modèle linéaire, mais non paramétrique, ne supposant pas une relation linéaire entre les deux types de données. Ainsi, les résidus des phénotypes ont été transformés en cinq variables indépendantes par fonction `ns()` du package R « splines » (R Core Team, 2013). Puis la régression a été effectuée par la fonction `lmFit()` du package R « limma » (Ritchie *et al.*, 2015) et la fonction `eBayes()` a permis de calculer les statistiques de tests du test de régression. Les valeurs-p obtenues ont été corrigées par la méthode de False Discovery Rate ($FDR < 0,05$). Les sondes significativement associées à au moins un phénotype ont été conservées et plus précisément analysées.

Ces dernières ont été annotées et une analyse d'enrichissement de fonctions ont été menées avec l'outil GOrilla (*Gene Ontology enRichment anaLysis and visualizAtion tool*). L'ensemble des gènes annotés par au moins une sonde sur la puce d'expression a été utilisé comme jeu global de données à partir duquel estimer les enrichissements de fonctions.



e) Approche multivariée pour l'identification de biomarqueurs

La méthode sPLS pour régression partielle des moindres carrés avec parcimonie (*sparse Partial Least Squares regression* ; Shen et Huang, 2008) a été appliquée pour intégrer les données d'expression des sondes précédemment sélectionnées et les données de variations phénotypiques, afin d'identifier des signatures de gènes pouvant prédire les phénotypes immunitaires. Nous avons utilisé l'implémentation de sPLS disponible dans le package mixOmics (Lê Cao *et al.*, 2009) sur R.

C. Résultats

1. Variabilité des phénotypes immunitaires mesurés

Au total, 55 caractères immunitaires mesurés à partir d'échantillons sanguins ont été inclus dans cette étude. Ils sont répartis en deux classes principales : 30 phénotypes descriptifs et 25 phénotypes de réponse (Tableau E3-9 et Annexe 2 pour la signification des abréviations). Les valeurs statistiques générales permettant d'estimer la variabilité de chaque caractère (minimum, maximum, moyenne, coefficient de variation) sont données dans le Tableau E3-9.

Tableau E3-9 : Statistiques générales des caractères immunitaires sanguins étudiés

Types	Caractères immunitaires	N	Femelles	Mâles	Valeurs brutes [min. - max.]	Moyenne	CV	Médiane
Phénotypes descriptifs	CRP	548	270	278	[729,5 - 1645000]	207384,5	91	146250
	HAPTO	549	271	278	[0,047 - 1,98]	0,4	97	0,3
	IgA	532	263	269	[111 - 4625]	831,6	57	740,0
	IgG_MH	546	269	277	[0,03657 - 3,791]	0,9	86	0,6
	NK	550	272	278	[0,2063 - 14,0808]	3,4	68	2,8
	NKT				[0,02417,83]	3,2	92	2,3
	gdT	308	150	158	[4,236 - 48,443]	29,1	26	28,7
	CD4+CD8+				[0,0498 - 7,40171]	2,3	49	2,1
	CD4+CD8 -				[0,007 - 18,60]	8,7	36	8,5
	CD4 - CD8+				[0,4554 - 16,5065]	7,7	37	7,5
	CD4 - CD8 -	550	272	278	[0,042 - 15,29]	0,8	122	0,7
	CD14+CD163+				[0,075 - 10,7]	1,8	94	1,3
	CD14+CD163 -				[0,16 - 9,59]	3,3	51	3,0
	cDC1				[0,003 - 4,028]	0,3	94	0,2
	cDC2				[0,002 - 2,57]	0,5	89	0,4
	CD14 - CD163+				[0,034 - 15,85]	3,3	68	2,9
	CD14 - CD163 -	471	233	238	[0,089 - 27,40]	5,5	84	4,0
	GB				[9,2 - 46,08]	17,7	23	17,2
	LYM				[6,2 - 45,5]	30,9	15	30,9
	MON				[2,2 - 10,9]	7,7	16	7,7
GRA	[44,7 - 91,6]				61,4	9	61,5	
EOS	[2,4 - 49,2]				6,2	69	5,2	
GR				[2,32 - 12,99]	6,6	11	6,6	



	Hgb				[3,84 - 23,81]	10,8	12	10,7
	Hct				[12,38 - 66,68]	35,2	11	35,1
	VGM				[41 - 67]	53,2	7	53,0
	TCMH				[12,09 - 32,29]	16,2	9	16,2
	CCMH				[28,23 - 48,3]	30,5	5	30,4
	PLT				[94 - 1144]	537,2	28	522,0
	GRA_LYM				[0,06769 - 1,01]	0,5	23	0,5
Phénotypes de réponse	IL1b_LPS_48h	529	261	268	[<0 - 68750]	4629,0	139	2985,0
	IL1b_PI_48h	540	268	272	[281,5 - 39700]	3304,7	103	2245,0
	IL1b_PIL_48h	517	258	259	[<0 - 27800]	7003,0	62	6025,0
	IL8_LPS_48h	432	206	226	[575,5 - 4940000]	198326,8	204	134750
	IL8_PI_48h	544	270	274	[10,95 - 352500]	31780,3	106	21750
	IL8_PIL_48h	536	268	267	[<0 - 114500]	23925,4	64	20450
	IL10_LPS_48h	470	235	235	[<0 - 2455]	471,7	61	423,5
	IL10_PI_48h	344	172	172	[<0 - 2350]	391,9	93	283,0
	IL10_PIL_48h	266	138	128	[43,4 - 3630]	728,7	82	266,0
	TNFa_LPS_48h	520	255	265	[<0 - 2700]	244,3	97	198,0
	TNFa_PI_48h	377	197	180	[134,5 - 23550]	3106,1	98	2070,0
	TNFa_PIL_48h	429	212	216	[<0 - 15150]	3327,0	66	2735,0
	IL2_4h	489	245	244	[<0 - 5018,84]	186,9	240	<0
	IL1b_4h	471	231	240	[<0 - 5227,03]	333,7	150	188,7
	IL8_4h	492	248	244	[<0 - 6367,91]	1054,7	92	766,6
	IL10_4h	524	261	263	[<0 - 33925,89]	2647,7	105	1753,4
	IL17_4h	522	262	260	[<0 - 1210,47]	225,4	67	187,6
	INFg_4h	515	256	259	[<0 - 15389,6]	111,4	655	<0
	TNFa_4h	492	248	244	[<0 - 4027,73]	218,0	235	<0
	Phagocytose				[16,69 - 71,64]	38,0	24	36,8
Phago_MFI				[5,99 - 8,281]	7,3	5	7,4	
Granulo_Phago				[46,3 - 93,41]	75,7	10	76,6	
Granulo_Phago_MFI	537	271	266	[6,2 - 8,504]	7,6	5	7,7	
Mono_Phago				[17,25 - 66,38]	40,1	22	40,3	
Mono_Phago_MFI				[5,002 - 7,44]	6,5	6	6,5	

CV = Coefficient de variation ou écart type relatif en % calculé selon la formule : $\frac{Ecart-type}{Moyenne} \times 100$

< 0 = Valeur phénotypique mesurée plus faible que la limite de détection de la technique utilisée

UA = Unité arbitraire

Les mesures de NFS permettent de confirmer que le groupe d'animaux utilisés dans cette étude sont « sains », sans signes cliniques d'infection. En effet, les valeurs en NFS sont globalement en accord avec les valeurs normales définies chez le porc. Ainsi, les lymphocytes (LYM) doivent être entre 35 et 54 % et nos valeurs s'étendent entre 6,2 % et 45,5 % avec une médiane à 30,9 %, quelques individus présentant un déficit en lymphocytes. Les leucocytes (GB) sont trouvés entre 11 et 22 $10^3/mm^3$. Pour cette mesure, deux individus sont largement au-dessus de la valeur normale (40,89 et 46,08 $10^3/mm^3$) mais la médiane est égale à 30,9 $10^3/mm^3$. Les éosinophiles doivent être entre 0 et 11 %, les globules rouges entre 5 et 8 %, le volume globulaire moyen (VGM) entre 50 et 68 fL, le TCMH entre 17 et 21 pg, le CCMH entre 27 et 40 g/dL, les plaquettes entre 250 et 750 $10^3/mm^3$.



Le coefficient de variation permet de distinguer des caractères immunitaires sanguins très stables dans la population et des caractères très variables d'un individu à un autre. Nous avons observé un coefficient de variation de moins de 10 % pour quatre caractères mesurés par NFS (VGM, TCMH, CCMH, GRA) et pour quelques mesures d'activités phagocytaires (Phago_MFI, Granulo_Phago, Granulo_Phago_MFI, Mono_Phago_MFI). Les lymphocytes T $\gamma\delta$ varient de 29 % entre nos individus. Parmi les phénotypes les plus variables au sein de cette population, en se basant sur les valeurs brutes de chaque mesure, nous avons noté que le taux d'interféron gamma 4h après stimulation par PIL présentait une variation de 655 %. Cela s'explique notamment par le fait que plus de la moitié des individus ont un taux en dessous du seuil de détection de la mesure en CBA (médiane = < 0). Nous avons observé le même phénomène pour les mesures de TNF alpha à 4h (CV = 235 % et médiane = < 0) et d'IL-2 à 4h (CV = 240% et médiane = < 0). D'autres phénotypes sont très variables au sein de la population étudiée comme le taux d'IL-8 48h après stimulation par le LPS, le taux d'IL-1 β à 4h (CV = 150%), le nombre de CD4-CD8- (CV = 122%), de cellules NKT (CV = 92%) et le taux de CRP (CV = 97%) et d'haptoglobine (CV = 91%) circulantes.

2. Facteurs biologiques et d'environnement qui influent sur la variabilité des phénotypes immunitaires mesurés

Nous avons systématiquement testé l'association des variations observées avec l'âge à la vaccination, la bande et le sexe. Le Tableau E3-10 présente les résultats de cette analyse.

Tableau E3-10 : Effets qui influent de manière significative sur les caractères immunitaires (code : 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1)**

Types de phénotypes immunitaires	Caractères immunitaires	Effet Sexe	Effet Bande	Effet Age Vaccination
Phénotypes descriptifs	CRP		*	**
	HAPTO	***	***	
	IgA	*	***	
	IgG_MH	**	***	*
	NK		***	
	NKT		***	
	gdT		***	
	CD4+CD8+		***	
	CD4+CD8 -	*	.	
	CD4 - CD8+		***	
	CD4 - CD8 -		***	
	CD14+CD163+	.	***	
	CD14+CD163 -	**	***	
	cDC1		***	
	cDC2		***	



	CD14 - CD163+	**	***	.
	CD14 - CD163 -	.	***	
	GB	**		*
	LYM		**	
	MON		***	
	GRA		**	
	EOS	**	***	
	GR		**	
	Hgb	**	***	
	Hct	***	.	
	VGM	***	*	
	TCMH	***	**	
	CCMH	*	***	
	PLT			
	GRA_LYM		**	
Phénotype de réponses fonctionnelles	IL1b_LPS_48h		*	
	IL1b_PI_48h		***	
	IL1b_PIL_48h	**	***	**
	IL8_LPS_48h		***	
	IL8_PI_48h		***	**
	IL8_PIL_48h		***	.
	IL10_LPS_48h		***	
	IL10_PI_48h		***	
	IL10_PIL_48h		***	
	TNFa_LPS_48h	*	***	
	TNFa_PI_48h		***	
	TNFa_PIL_48h	**	***	
	IL2_4h		*	
	IL1b_4h	**	***	
	IL8_4h		***	
	IL10_4h		***	
	IL17_4h		***	
	INFg_4h		.	
	TNFa_4h		***	
	Phagocytose	***	***	
	Phago_MFI	***	***	
	Granulo_Phago	***	***	
	Granulo_Phago_MFI	**	***	
	Mono_Phago		***	
	Mono_Phago_MFI		***	

Les porcs sont élevés par bandes. Cette conduite d'élevage consiste à remplir en une seule fois un bâtiment d'élevage avec des animaux de même âge. Ainsi, les porcs d'une même « Bande » sont contemporains ; ils ont été élevés en même temps et dans les mêmes conditions d'environnement. Tous les caractères immunitaires sanguins sont associés significativement avec la bande sauf le taux de plaquettes, le



taux de globules blancs. La bande a été donc gardée comme un effet fixe dans les modèles d'analyse de tous les phénotypes.

Les porcelets ne naissent pas tous en même temps mais, en pratique, ils sont vaccinés le même jour, à environ 35 jours de vie ici dans le projet SUS_FLORA, plus ou moins quelques jours en fonction de leur date « réelle » de naissance. Comme attendu, la variation du taux d'IgG anti-*Mycoplasma hyopneumoniae* est influencée par l'âge à la vaccination. De manière plus surprenante, l'âge semble avoir un effet également sur quelques mesures de cytokines après stimulation (IL-8, IL-1 β), sur la variation du taux de CRP circulante et sur le taux de globules blancs. L'âge à la vaccination a été considéré comme une covariable linéaire.

Le sexe des porcelets semble influencer plutôt les phénotypes descriptifs que les phénotypes de réponse, et plus particulièrement les comptages cellulaires, les anticorps circulants et l'haptoglobine. Parmi les phénotypes de réponse, le sexe influence quatre paramètres dont la phagocytose et les taux d'IL-1 β et de TNF- α après stimulation *in vitro*. Le sexe a été considéré comme un effet fixe.

3. Covariations des caractères immunitaires sanguins

a) Corrélations et classification hiérarchique

Nous avons analysé les covariations entre les caractères par le calcul des corrélations de Spearman. Parmi les 1485 relations possibles, 174 étaient significatives ($FDR \leq 0,05$). Ces corrélations sont représentées à la Figure E3-33. Sur la base des corrélations, nous avons établi une classification hiérarchique non supervisée des données de mesures des caractères immunitaires sanguins après correction pour le sexe, l'âge à la vaccination, la bande, la loge, la mère et les facteurs inhérents à chaque type de mesures (Figure E3-33 et E3-34).

Les taux d'anticorps circulants IgA et IgG anti-*Mycoplasma hyopneumoniae* ainsi que le taux d'INF- γ 4h après stimulation PIL ne sont significativement corrélés à aucun caractère immunitaire. Les taux d'haptoglobine et de CRP circulantes, deux protéines produites par les hépatocytes, sont significativement et positivement corrélés ($\rho = 0,32$). Elles sont également toutes deux positivement corrélées à l'efficacité de phagocytose (Phago_MFI ; respectivement $\rho = 0,32$ et $\rho = 0,19$) et seule l'haptoglobine est corrélée à l'efficacité de phagocytose des granulocytes (Phago_Granulo_MFI ; $\rho = 0,21$).



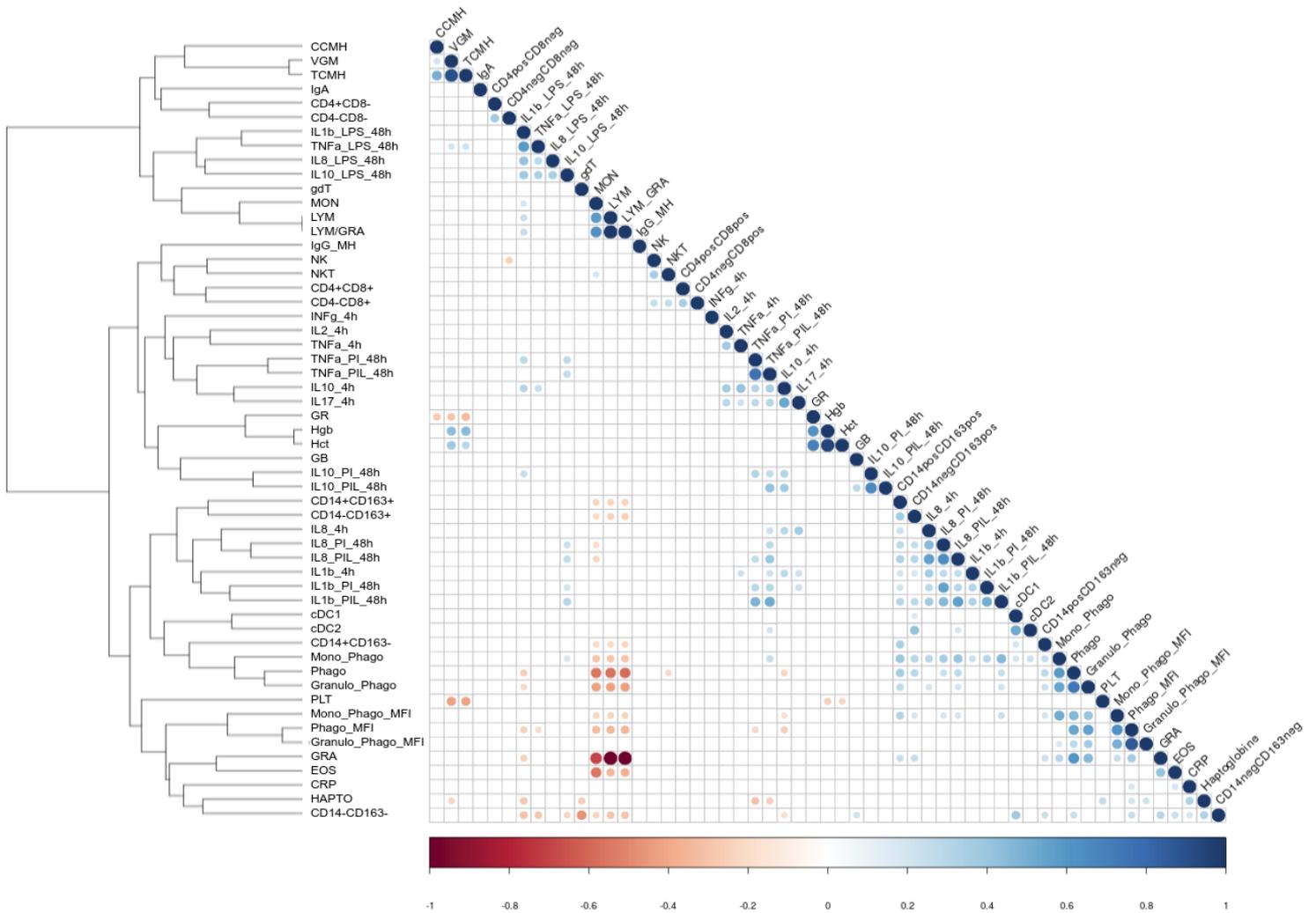


Figure E3-33 : Matrice des corrélations (Spearman) entre les 55 caractères immunitaires (Corrplot). Seules les corrélations significatives après correction (FDR < 0,05) sont représentées. Les caractères sont triés en fonction de la classification hiérarchique (Figure E3-34).

Les taux de cellules NKT sont significativement et négativement corrélés à la phagocytose ($\rho = -0,22$). Des caractères de phagocytose sont également négativement corrélés avec la mesure d'IL-10 4h après stimulation ($\rho = -0,23$ en moyenne). Cela peut s'expliquer par la fonction régulatrice de l'IL-10 sur des cytokines qui activent/attirent les cellules phagocytaires, comme l'IL-8 trouvée corrélée positivement avec les caractères de phagocytose ($0,24 > \rho > 0,41$) ou produites par les cellules phagocytaires comme l'IL-1 β . Il est intéressant de noter que la production d'IL-1 β , cytokine produite par les macrophages, 48h après stimulation au LPS, est négativement corrélée au pourcentage de cellules phagocytaires ($\rho = -0,25$), à la quantité de bactéries phagocytées ou efficacité phagocytaire ($\rho = -0,25$) et au pourcentage de granulocytes phagocytaires ($\rho = -0,21$). Le taux de la même cytokine mesuré 48h après stimulation PI et PIL est corrélé positivement avec la proportion de monocytes capables de phagocyter ($\rho = 0,31$ et $\rho = 0,45$, respectivement).



La proportion de cellules capables de phagocyter, la quantité de bactéries phagocytées et plus spécifiquement la proportion de granulocytes capables de phagocyter après stimulation sont fortement positivement et significativement corrélées au pourcentage de granulocytes quantifiés par NFS ($0,36 < \rho < 0,61$). Donc plus il y a de granulocytes circulants, plus la proportion de granulocytes qui vont phagocyter en cas de stimulation va être importante, mais cela ne sera pas corrélé au nombre de bactéries phagocytées par granulocyte. Ces trois caractères de phagocytose sont négativement corrélés à la proportion de monocytes et de lymphocytes mesurés en NFS. Il est intéressant de noter qu'il n'y a pas de corrélation évidente entre la proportion de monocytes mesurée en NFS et les caractères phagocytaires spécifiques des monocytes. La proportion de monocytes capables de phagocyter est également corrélée au taux de cellules dendritiques (cDC1 et cDC2). Enfin, plusieurs caractères phagocytaires sont également corrélés au pourcentage des monocytes CD14+ CD163+ (III), CD14+ CD163- (I), CD14- CD163+ (IV) et CD14- CD163- (II).

Il est intéressant d'observer que pour les cytokines IL-1 β , IL-8 et IL-10, mesurées en CBA 4h après stimulation PIL et en ELISA 48h après stimulation PI et PIL, le taux de production est corrélé positivement alors que le taux de ces cytokines mesurées 48h après stimulation par LPS n'est pas corrélé. Ainsi, pour ces trois cytokines, les mesures à 4h et 48h pour la même stimulation sont corrélées, sauf pour TNF- α . Cela peut s'expliquer par l'implication du TNF- α dans les réactions aiguës inflammatoires, avec peut-être une cinétique différente des autres cytokines.

Il semble également que la stimulation par le LPS qui mime une infection bactérienne (Gram-) entraîne des taux de réponses cytokiniques différents de ceux obtenus après stimulation PI, qui déclenche une réponse immunitaire non spécifique, ou PIL, qui combine les deux stimulations précédentes. Cela se confirme par une absence de corrélation entre la stimulation LPS et PI/PIL pour les mesures à 48h des cytokines IL-1 β , IL-8, IL-10 et TNF- α , alors que les taux 48h après PI et PIL sont fortement et positivement corrélés ($0,51 < \rho < 0,76$).

Quarante-huit heures après stimulation LPS, toutes les cytokines mesurées sont corrélées entre elles. Certaines sont même fortement corrélées comme IL-1 β et TNF- α ($\rho = 0,59$) qui activent toutes deux l'apoptose ou encore IL-1 β et IL-8 ($\rho = 0,42$), toutes deux liées à des mécanismes de défense par phagocytose. A 48h après stimulation PI, seules IL-1 β /IL-8 ($\rho = 0,56$) et IL-10/TNF- α sont significativement et positivement corrélées. Cependant, 4h après stimulation PIL, l'IL-10, protéine régulant et inhibant les autres cytokines, est positivement corrélée à toutes les cytokines mesurées ($0,29 < \rho < 0,54$). Cela peut s'expliquer par le temps très précoce de mesure, auquel l'action inhibitrice de l'IL-10 n'a pas encore fait effet sur les taux des autres cytokines. L'IL-17, cytokine produite par les lymphocytes Th17, est aussi corrélée à l'IL-1 β , ($\rho = 0,29$), l'IL-2 ($\rho = 0,29$), et l'IL-8 ($\rho = 0,37$).



Les taux de cellules NK et NKT sont corrélés positivement ($\rho = 0,35$). Comme mentionné précédemment, les proportions de cellules NKT sont également négativement corrélées à celles des cellules capables de phagocyter ($\rho = -0,22$), et positivement corrélées au taux de monocytes mesurés en NFS ($\rho = 0,19$) et au pourcentage de lymphocytes T cytotoxiques (CD4- CD8+) ($\rho = 0,26$). Les cellules NK sont également corrélées positivement au taux de lymphocytes CD4- CD8+ ($\rho = 0,25$) mais négativement corrélées aux CD4- CD8- ($\rho = -0,25$).

On peut également noter que le taux de lymphocytes $T\gamma\delta$, nombreux chez le porc, est corrélé au taux d'haptoglobine et de cellules CD14-CD163-. De plus, les cellules dendritiques cDC1 et cDC2 sont fortement corrélées ($\rho = 0,52$) mais les cDC1 sont corrélées aux CD14- CD163- ($\rho = 0,38$) et les cDC2 le sont aux CD14-CD163+ ($\rho = 0,41$). Enfin les CD4- CD8+ sont corrélées aux CD4+ CD8+ ($\rho = 0,35$) et les CD4- CD8- sont corrélées aux CD4+ CD8- ($\rho = 0,37$).

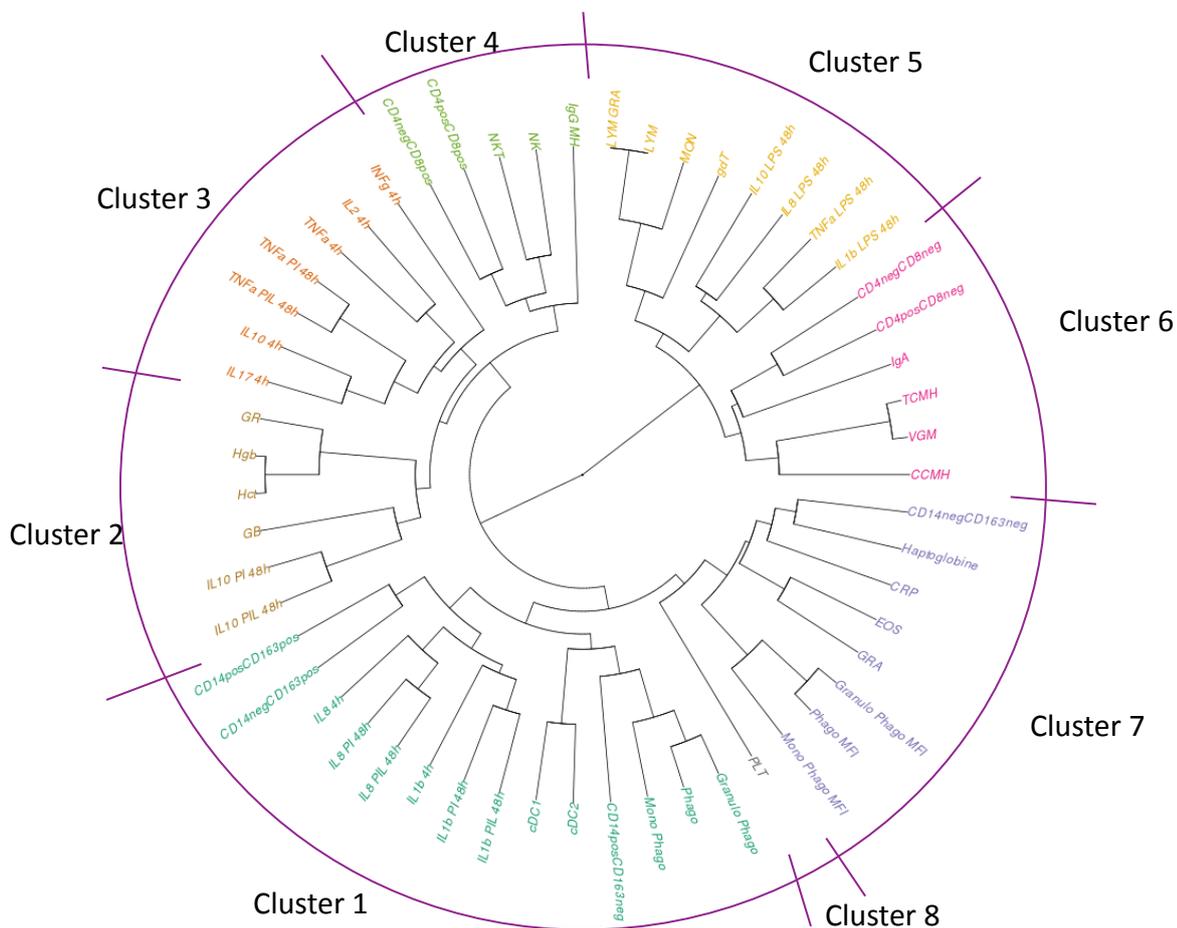


Figure E3-34 : Classification hiérarchique non supervisée répartissant les caractères immunitaires en 8 groupes. Pour plus de faciliter d'interprétation, nous avons attribué des numéros de 1 à 8 aux sous-groupes de caractères.

La classification hiérarchique (Figure E3-34) nous permet d'observer que le taux de plaquettes dans le sang (PLT) ne se groupe avec aucun autre caractère immunitaire. Les trois caractères TCMH, VGM et CCMH se regroupent en un cluster. Le taux de cellules NK et NKT se regroupent avec les cellules CD4+CD8+ et CD4-



CD8+. Trois caractères de phagocytose se regroupent dans un même ensemble (cluster 7) : Granulo_Phago_MFI, Phago_MFI et Mono_Phago_MFI. Les trois autres caractères de phagocytoses sont regroupés dans un autres sous ensemble (cluster 1). Les globules rouges (GR) l'hématocrite (Hct) et l'hémoglobine (Hgb) sont également proches dans cette classification. Les mesures de cytokines sont réparties dans trois clusters différents. On remarque tout de même que les mesures du TNF- α 48h après stimulation PI et PIL sont très proches, et à proximité de la mesure du TNF- α 4h après stimulation PIL. On retrouve la même organisation pour l'IL-8 et l'IL-1 β . Les mesures après stimulation au LPS sont quant à elles regroupées ensemble (cluster 5). Les lymphocytes et monocytes mesurés par NFS sont dans le même sous ensemble que les lymphocytes T $\gamma\delta$.

b) Corrélation directe ou partielle et analyse de réseau

Afin d'identifier les corrélations entre les phénotypes les plus informatives, nous avons appliqué l'algorithme de corrélation partielle (PCIT) qui permet d'éliminer les relations de corrélations indirectes entre deux phénotypes médiées par une corrélation à un même troisième phénotype. L'analyse des corrélations entre les caractères immunitaires a ainsi mis en évidence 242 relations directes significatives parmi les corrélations identifiées (Figure E3-33). La fréquence et la valeur de ces corrélations sont représentées sur la Figure E3-35.

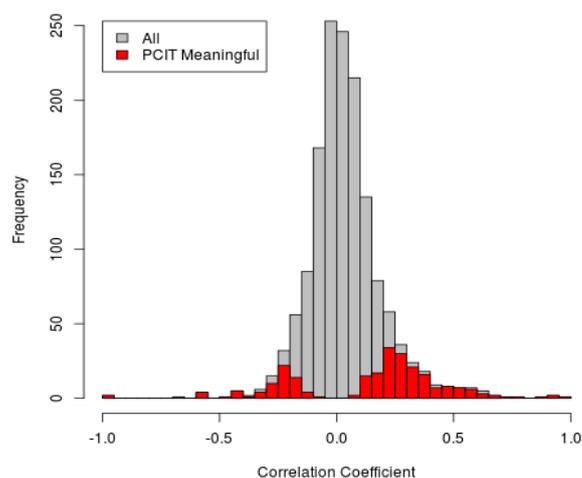


Figure E3-35 : Distribution des coefficient de corrélations. Les corrélations directes significatives déterminées par PCIT sont représentées en rouge



Le réseau des corrélations significatives est représenté Figure E3-36. Sur ce réseau, plus le disque du phénotype est large, plus le phénotype est corrélé à d'autres et donc majeur au sein du réseau. L'épaisseur des liens est proportionnelle à la valeur de la corrélation. Plus un phénotype est éloigné du centre du réseau, plus sa variation sera indépendante des autres phénotypes. Ainsi, sur ce réseau, on retrouve les anticorps circulants IgA ou les cellules CD4- CD8- et CD4+ CD8-, reflétant le peu de corrélation entre ces deux caractères et les autres. De manière globale, il semble qu'il y a moins d'interactions entre phénotypes descriptifs et de réponse qu'au sein de chaque groupe de phénotypes. Pour affiner ces résultats, il sera important de normaliser les données de réponses avec la quantité de leucocytes présents dans les volumes de sang utilisés pour les dosages ex vivo.

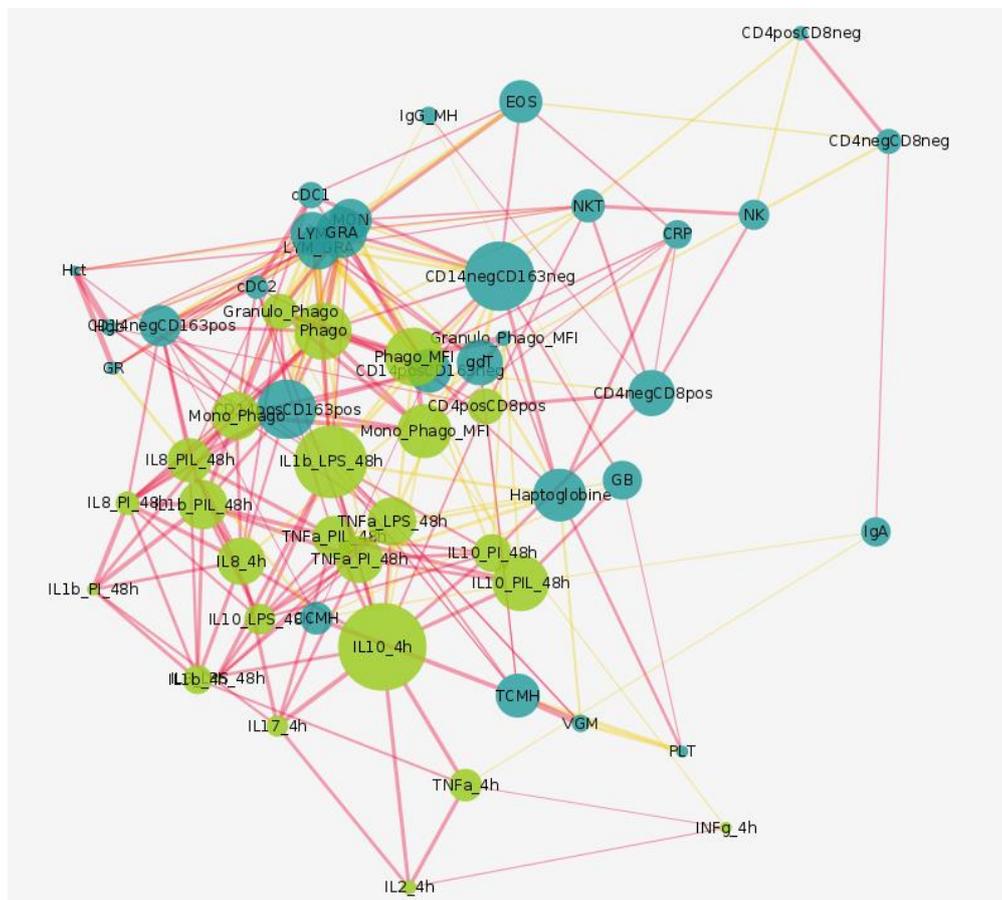


Figure E3-36 : Représentation du réseau des corrélations directes significatives entre les caractères immunitaires, déterminées par PCIT. Les phénotypes descriptifs sont symbolisés par des ronds bleus et les phénotypes de réponse par des ronds verts. Les corrélations positives sont signifiées par des traits rouges et les négatives par des traits jaunes. La taille des ronds est proportionnelle au nombre de corrélations avec le phénotype qu'il représente



4. Héritabilité des caractères immunitaires sanguins

L'héritabilité de la variation de 18 caractères immunitaires a été estimée en tenant compte des effets susceptibles d'interférer avec sa mesure (Tableau E3-8) et les valeurs estimées sont présentées dans la Figure E3-37. Nous avons estimé une héritabilité significative modérée pour le taux d'haptoglobine circulant ($h^2 = 0,30$), le taux d'IL-10 mesuré 4h après stimulation PIL ($h^2 = 0,27-0,37$) et le taux de TNF- α ($h^2 = 0,32$) et d'IL-8 mesurés 48h après stimulation PIL ($h^2 = 0,29$). Nous avons également estimé une forte héritabilité pour le TNF- α mesuré 48h après stimulation PI ($h^2 = 0,48$).

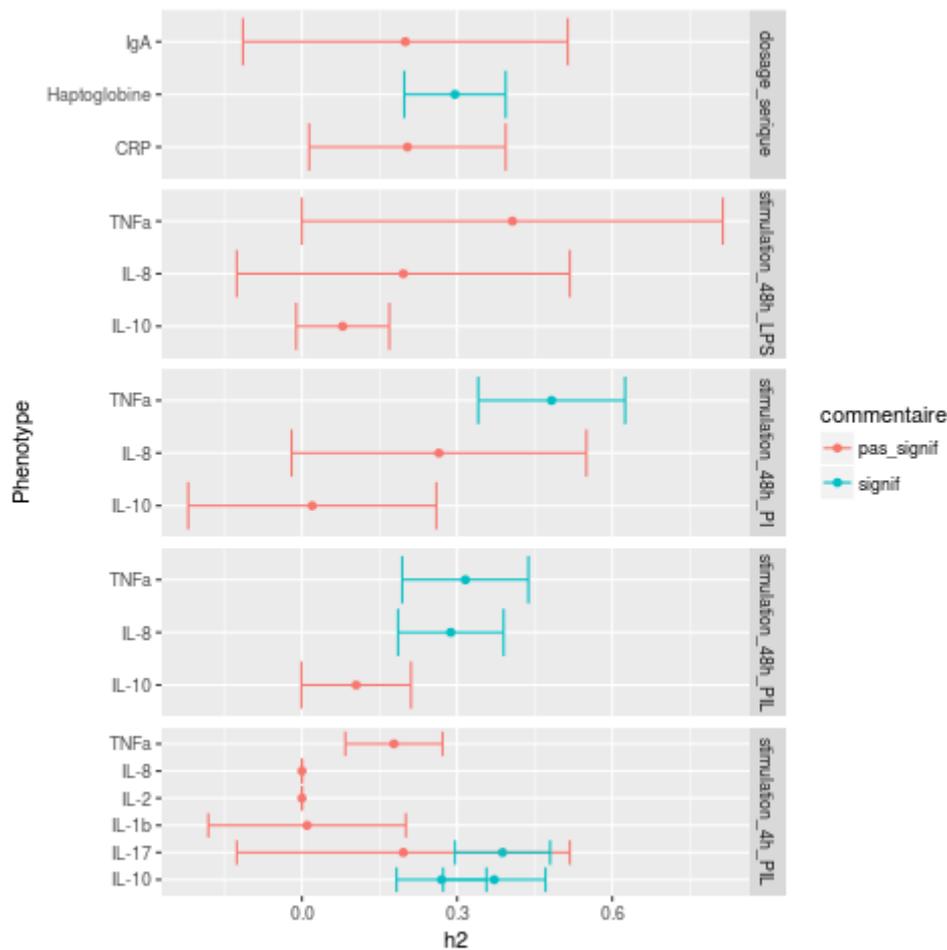


Figure E3-37 : Estimation de l'héritabilité et de l'écart-type correspondant pour 18 caractères immunitaires.

5. Identification de biomarqueurs à partir du transcriptome sanguin

- a) Covariations entre niveau d'expression des gènes et caractères immunitaires



Dans une première étape, nous avons identifié, pour chaque paramètre immunitaire traité indépendamment, une liste de sondes dont la variation du niveau d'expression entre les individus est associée à la variation du phénotype. Globalement, nous avons identifié 6 094 sondes associées à au moins un caractère immunitaire sanguin, dont 4 929 sont annotées pour 3 504 gènes porcins. Le détail du nombre de sondes et gènes pour l'ensemble des caractères immunitaires ou pour les sous-groupes des paramètres descriptifs ou de réponse est résumé dans les Tableaux E3-11 et Tableau Supplémentaire 1. Cet ensemble de gènes est significativement enrichi pour des fonctions majoritairement liées à l'immunité.

Pris séparément, les gènes associés à des caractères descriptifs (n=1 605 gènes) et les gènes associés à des caractères de réponse fonctionnelle (n=2 760 gènes) sont également enrichis pour des fonctions majoritairement impliquées dans l'immunité. Les 20 premières fonctions les plus enrichies pour chaque analyse sont présentées dans les Tableaux E3-12 à E3-14.

Tableau E3-11 – Nombre de sondes et gènes dont le niveau d'expression a été trouvé associé à la variation des phénotypes

	Phenotypes	Animaux phénotypés	Femelles	Mâles	Sondes associées (FDR < 0,05)	Gènes associés (FDR < 0,05)
Phénotypes de réponse	IL2_4h	210	99	111	2	1
	IL1b_4h	206	96	110	49	32
	IL8_4h	205	96	109	21	14
	IL10_4h	221	102	119	4	3
	IL17_4h	222	103	119	13	11
	INFg_4h	214	99	115	389	279
	TNFa_4h	205	95	110	0	0
	IL1b_LPS_48h	230	105	125	147	94
	IL1b_PI_48h	240	111	129	17	10
	IL1b_PIL_48h	232	108	124	22	18
	IL8_LPS_48h	194	86	108	342	241
	IL8_PI_48h	241	112	129	128	86
	IL8_PIL_48h	236	110	123	247	165
	IL10_LPS_48h	205	96	109	74	50
	IL10_PI_48h	141	62	79	21	16
	IL10_PIL_48h	110	54	56	5	4
	TNFa_LPS_48h	231	105	126	4	3
	TNFa_PI_48h	188	92	96	56	34
	TNFa_PIL_48h	185	86	99	114	82
	Phago				2147	1201
Phago_MFI				266	134	
Granulo_Phago_MFI	234	110	124	20	14	
Mono_Phago_MFI				516	379	
Granulo_Phago				965	459	
Mono_Phago				867	572	
Phénotypes descriptifs	gdT	162	69	93	96	66
	CD4+CD8+				13	9
	CD4+CD8-				41	30
	CD4-CD8+				79	43
	CD4-CD8-				40	24



	NK				177	94
	NKT				217	126
	CD14+CD163+				272	164
	CD14+CD163-				32	27
	cDC1	243	112	131	3	2
	cDC2				3	2
	CD14-CD163+				61	45
	CD14-CD163-				379	224
	GB				22	15
	LYM				803	374
	MON				966	550
	GRA				947	425
	EOS				28	16
	GR				28	16
	Hgb	195	87	108	57	38
	Hct				39	24
	VGM				21	10
	TCMH				54	37
	CCMH				27	18
	PLT				5	2
	LYM_GRA				887	409
	CRP	242	112	130	265	177
	Haptoglobine	243	112	131	2	2
	IgA	235	110	125	38	28
	IgG_MH	240	110	130	56	35

¹ Etude 1, Maroilley *et al.* (en révision)

Tableau E3-12 – Top 20 des fonctions biologiques trouvées enrichies sur la base les gènes dont la variation d'expression dans les sang a été trouvée associée à la variation d'au moins un paramètre immunitaire descriptif ou de réponse

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0002252	immune effector process	1.14E-9	1.68E-5	1.40
GO:0002274	myeloid leukocyte activation	1.71E-8	1.26E-4	1.47
GO:0006887	exocytosis	1.68E-7	8.28E-4	1.38
GO:0002376	immune system process	2.17E-7	8.02E-4	1.21
GO:0045055	regulated exocytosis	4.31E-7	1.27E-3	1.38
GO:0002263	cell activation involved in immune response	4.79E-7	1.18E-3	1.40
GO:0036230	granulocyte activation	5.7E-7	1.2E-3	1.44
GO:0042119	neutrophil activation	7.22E-7	1.33E-3	1.44
GO:0002366	leukocyte activation involved in immune response	7.49E-7	1.23E-3	1.40
GO:0002275	myeloid cell activation involved in immune response	8.16E-7	1.2E-3	1.43
GO:0043299	leukocyte degranulation	1.78E-6	2.38E-3	1.42
GO:0043312	neutrophil degranulation	1.84E-6	2.27E-3	1.43
GO:0002283	neutrophil activation involved in immune response	2.14E-6	2.43E-3	1.42
GO:0010941	regulation of cell death	2.72E-6	2.87E-3	1.22
GO:0043067	regulation of programmed cell death	7.54E-6	7.42E-3	1.21
GO:0051246	regulation of protein metabolic process	1.45E-5	1.33E-2	1.15
GO:0046903	secretion	1.45E-5	1.26E-2	1.26
GO:0048519	negative regulation of biological process	1.49E-5	1.22E-2	1.10
GO:0042981	regulation of apoptotic process	1.93E-5	1.5E-2	1.21



GO:0006796 phosphate-containing compound metabolic process 1.96E-5 1.44E-2 1.17

Tableau E3-13– Top 20 des fonctions biologiques trouvées enrichies sur la base les gènes dont la variation d’expression dans les sang a été trouvée associée à la variation d’au moins un paramètre immunitaire descriptif

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0002376	immune system process	7.13E-19	1.05E-14	1.61
GO:0006952	defense response	1.27E-17	9.38E-14	1.92
GO:0002252	immune effector process	3.42E-17	1.68E-13	1.93
GO:0006955	immune response	3.45E-15	1.27E-11	1.91
GO:0002682	regulation of immune system process	1.44E-13	4.24E-10	1.62
GO:0002274	myeloid leukocyte activation	2.57E-13	6.32E-10	2.03
GO:0006954	inflammatory response	1.97E-12	4.15E-9	2.18
GO:0006887	exocytosis	1.72E-11	3.16E-8	1.81
GO:0045055	regulated exocytosis	2.23E-11	3.66E-8	1.85
GO:0002684	positive regulation of immune system process	2.84E-11	4.19E-8	1.71
GO:0002263	cell activation involved in immune response	2.89E-11	3.87E-8	1.89
GO:0048583	regulation of response to stimulus	4.14E-11	5.09E-8	1.29
GO:0080134	regulation of response to stress	5.69E-11	6.45E-8	1.55
GO:0001775	cell activation	6.78E-11	7.14E-8	1.66
GO:0009605	response to external stimulus	7.3E-11	7.18E-8	1.58
GO:0002366	leukocyte activation involved in immune response	8.8E-11	8.11E-8	1.87
GO:0006950	response to stress	9.2E-11	7.98E-8	1.35
GO:0043299	leukocyte degranulation	1.51E-10	1.23E-7	1.95
GO:0042119	neutrophil activation	1.59E-10	1.23E-7	1.96
GO:0045321	leukocyte activation	1.76E-10	1.3E-7	1.70

Tableau E3-14– Top 20 des fonctions biologiques trouvées enrichies sur la base les gènes dont la variation d’expression dans les sang a été trouvée associée à la variation d’au moins un paramètre immunitaire de réponse

GO Term	Description	P-value	FDR q-value	Enrichment
GO:0002376	immune system process	7.63E-10	1.12E-5	1.29
GO:0002252	immune effector process	8.71E-9	6.43E-5	1.43
GO:0002682	regulation of immune system process	1.54E-8	7.55E-5	1.33
GO:0002684	positive regulation of immune system process	6.55E-8	2.42E-4	1.40
GO:0002274	myeloid leukocyte activation	4.69E-7	1.38E-3	1.48
GO:0044237	cellular metabolic process	6.5E-7	1.6E-3	1.08
GO:0006887	exocytosis	9.6E-7	2.02E-3	1.40
GO:0001817	regulation of cytokine production	1.46E-6	2.69E-3	1.43
GO:0036230	granulocyte activation	1.62E-6	2.66E-3	1.49
GO:0002263	cell activation involved in immune response	1.95E-6	2.87E-3	1.44
GO:0002366	leukocyte activation involved in immune response	2.03E-6	2.72E-3	1.44
GO:0042119	neutrophil activation	2.18E-6	2.68E-3	1.48
GO:0050776	regulation of immune response	3.52E-6	3.99E-3	1.34
GO:0001819	positive regulation of cytokine production	4.34E-6	4.57E-3	1.52
GO:0002275	myeloid cell activation involved in immune response	4.84E-6	4.75E-3	1.45
GO:0045055	regulated exocytosis	5.01E-6	4.62E-3	1.39
GO:0006796	phosphate-containing compound metabolic process	5.36E-6	4.65E-3	1.21
GO:0071704	organic substance metabolic process	7.54E-6	6.18E-3	1.07
GO:0045321	leukocyte activation	7.85E-6	6.1E-3	1.33
GO:0044238	primary metabolic process	9.84E-6	7.25E-3	1.07



6. Identification de biomarqueurs candidats

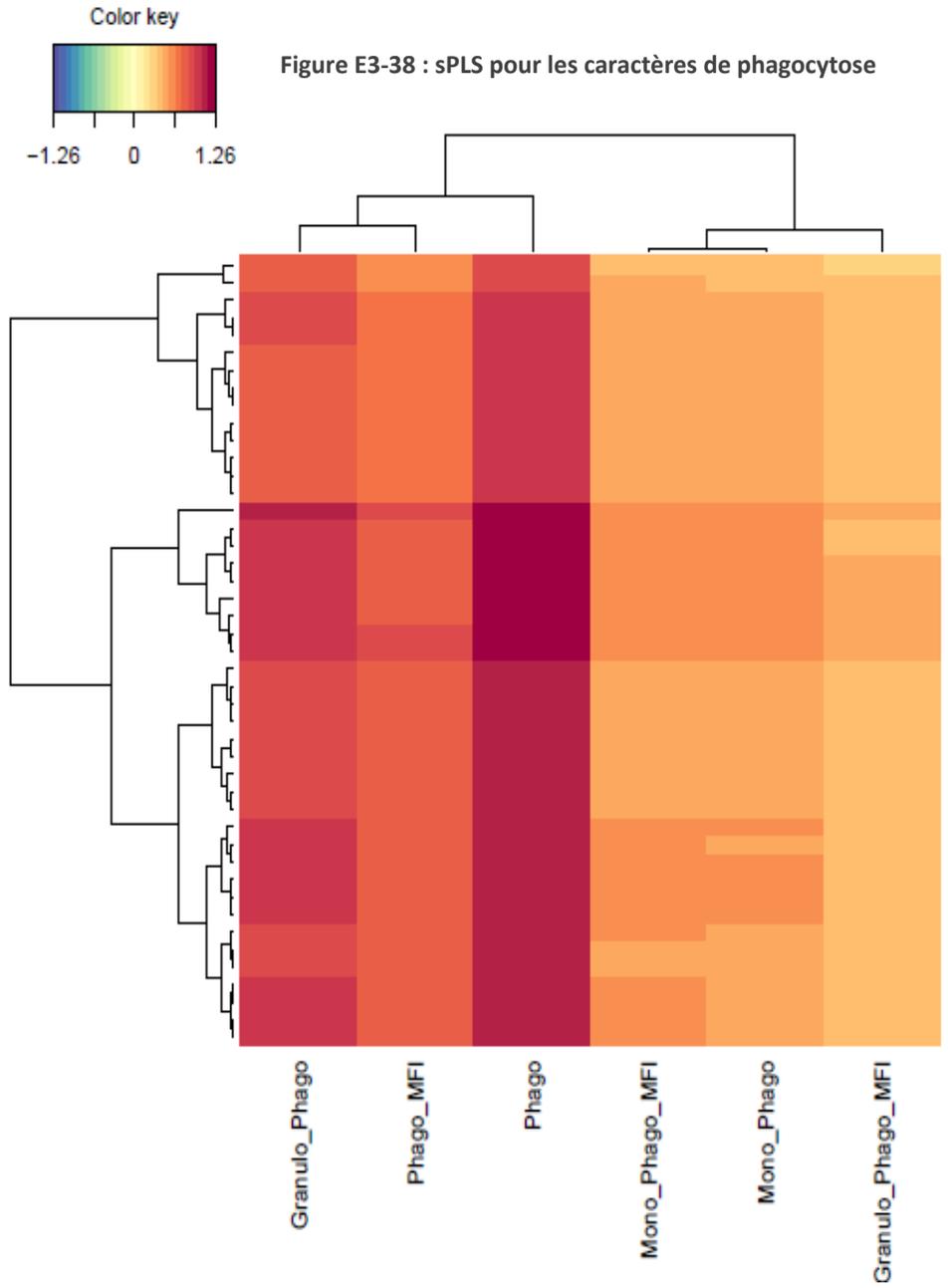
Dans une précédente étude, Mach *et al.* (2013) ont pu identifier une liste de gènes biomarqueurs candidats : *RALGDS* pour la production d'IL-2 ; *GATM*, *SACRB1* et *ALOX12* pour la capacité de phagocytose ; *GNLY*, *FASLG*, *DDIT4*, *GZMB*, *CTSG1*, *KLRG1* et *CX3CR1* pour la quantité de cellules CD4-CD8+. Nous avons pu retrouver le gène *GATM* associé au nombre de granulocytes détectés par NFS (GRA), au ratio LYM_GRA et à la quantité de CD4+CD8-. Nous avons également confirmé *SCARB1* comme un biomarqueur candidat intéressant pour le nombre de granulocytes détectés par NFS (GRA), le taux d'IL-8 produit in vitro après stimulation au LPS (IL8_48h_LPS), au nombre de cellules phagocytaires (Phago) et à la quantité de cellules CD14+CD163+, le stade III de la maturation des monocytes. Nous avons également identifié *DDIT4* associé à la quantité d'IFN- γ produit 4h après stimulation in vitro, *GZMB* associé au taux de cellules NK, et *CX3CR1* associé au taux de cellules NK, NKT et CD4-CD8+. Sur 11 biomarqueurs repérés par Mach *et al.* (2013), cinq sont retrouvés dans notre étude, avec deux d'entre eux associés à la variation des mêmes phénotypes (*SCARB1* pour la phagocytose et *CX3CR1* pour les taux de cellules T CD4-CD8+).

Afin d'identifier des biomarqueurs associés à plusieurs phénotypes, nous avons poursuivi l'analyse en nous basant sur les 4 926 sondes annotées pour 3 504 gènes. Nous avons mené une analyse sPLS afin d'intégrer la variation des caractères immunitaires et la variation de l'expression de ces sondes. Il s'agit d'identifier par régression si les gènes permettent de prédire la variation des caractères immunitaires. Pour toutes les analyses, nous avons fixé à 45 le nombre de sondes/gènes, correspondant au nombre de gènes pouvant être inclus dans une puce Fluidigm.

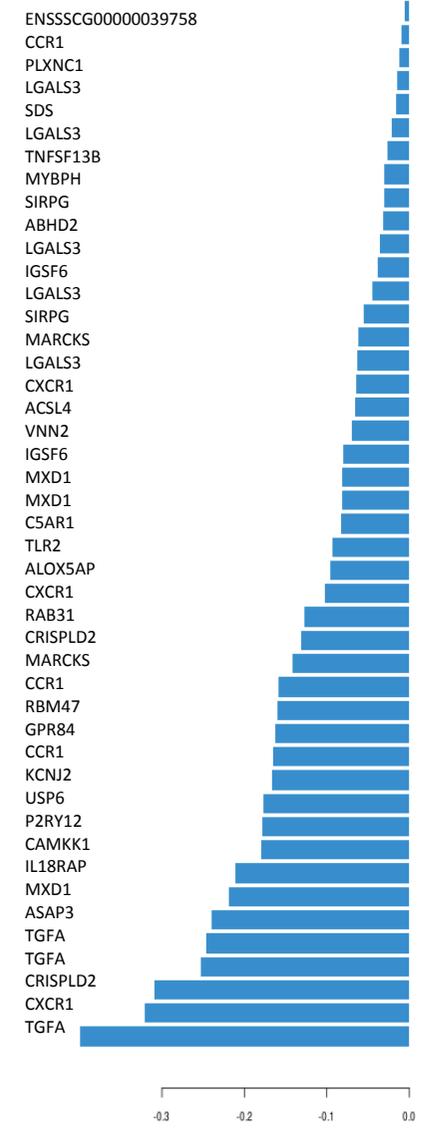
La première analyse que nous avons menée incluait tous les caractères immunitaires. Aucune signature de biomarqueurs sanguins significative n'a pu être mise en évidence ($Q_2 = 0.00134$). Nous avons alors tenté d'identifier des biomarqueurs pour des groupes de caractères immunitaires constitués a priori, sur la base d'informations biologiques qui nous semblaient pertinentes. Par exemple, pour les six caractères de capacité phagocytaire, nous avons pu identifier un ensemble de gènes prédictifs de leur variation ($Q_2 = 0.190$) (Figure E3-38). En revanche, pour des caractères comme les mesures d'IL-8 ($Q_2 = 0.012$) ou IL-1b ($Q_2 = 0.041$), aucun groupe de gènes significatifs n'a pu être mis en évidence malgré des tendances. Nous avons fait le même constat pour les comptages cellulaires CD4+/- et CD8+/- ($Q_2 = 0.034$).



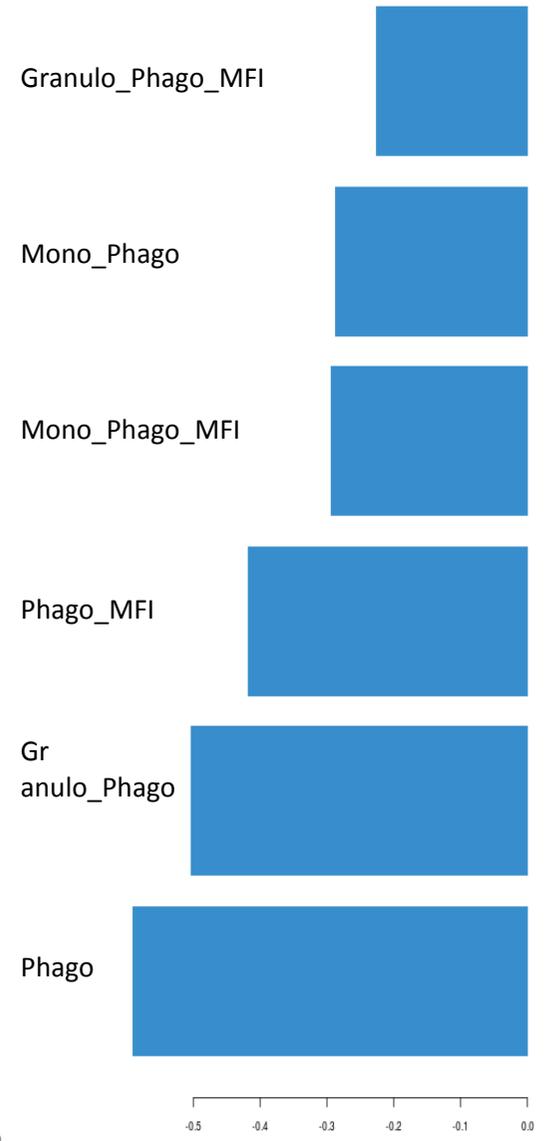
Figure E3-38 : sPLS pour les caractères de phagocytose



Loadings on comp 1 Block 'X'



Loadings on comp 1 Block 'Y'



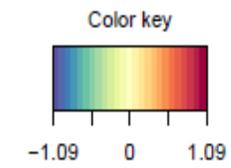
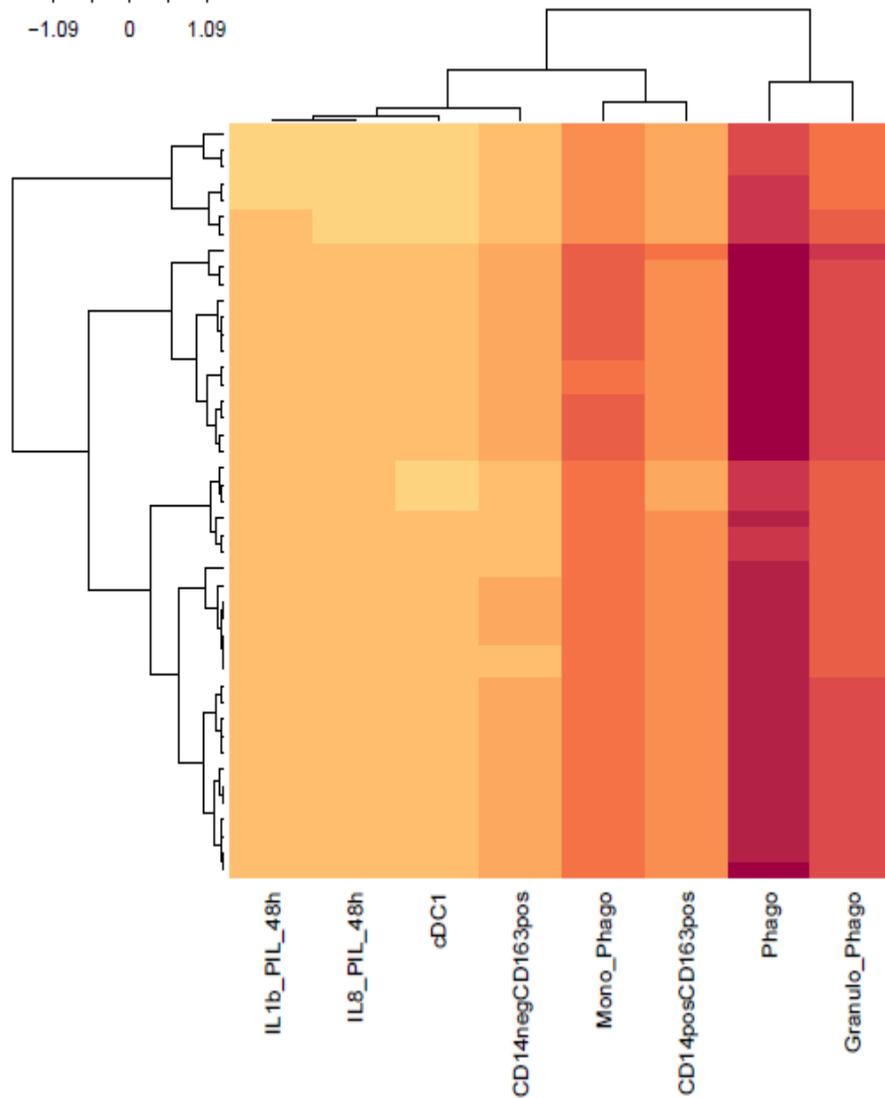
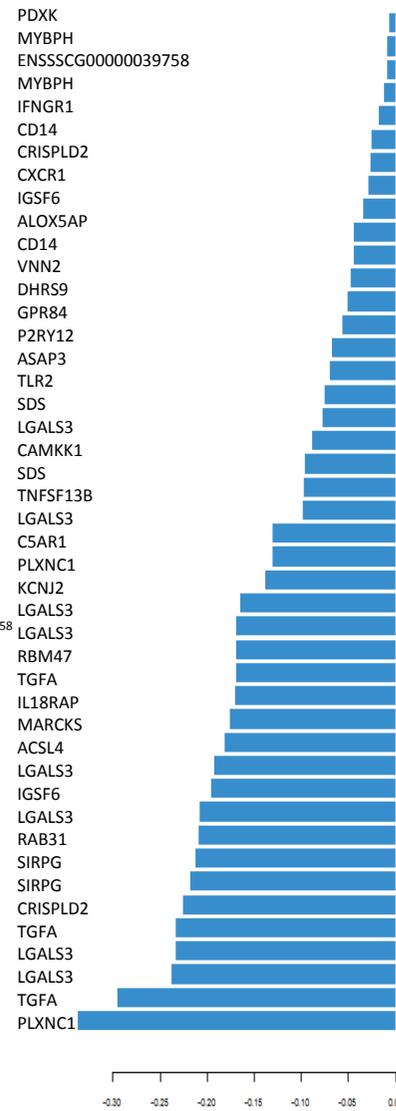


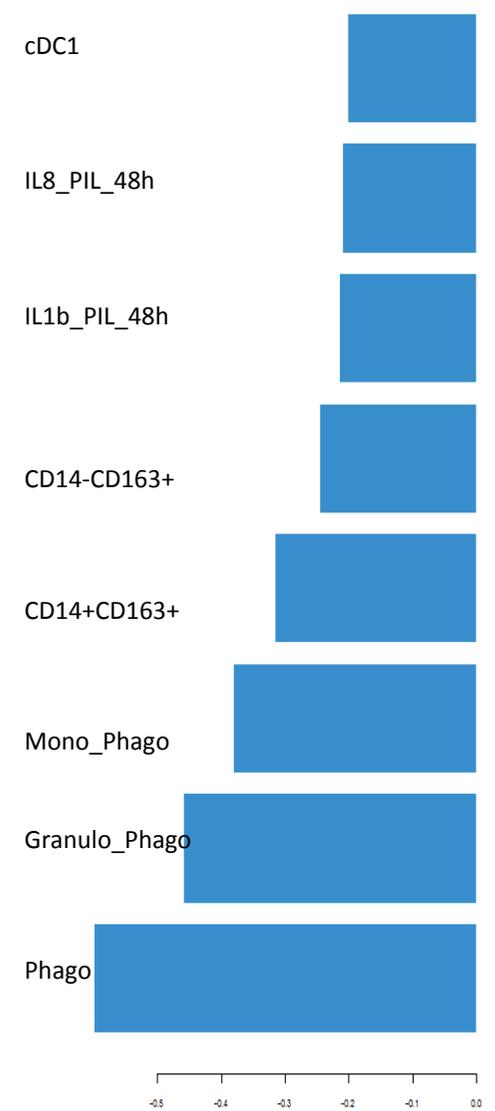
Figure E3-39 : sPLSA pour une partie des caractères du cluster 1



Loadings on comp 1
Block 'X'



Loadings on comp 1
Block 'Y'



Nous avons ensuite cherché à identifier des biomarqueurs pour des caractères regroupés selon la classification hiérarchique non supervisée établie lors de l'étude des covariations entre paramètres immunitaires (Figure E3-34). Ainsi, nous nous sommes intéressés au cluster 1 qui inclue 14 caractères descriptifs et de réponse cytokiniques et cellulaires liés à la phagocytose (Figure E3-34). La première analyse sPLS a révélé un ensemble de gènes non significatifs ($Q_2 = 0.0636$). Nous avons alors restreint l'analyse à un sous-groupe de huit phénotypes : CD14posCD163pos, CD14negCD163pos, IL8_PIL_48h, IL1b_PIL_48h, cDC1, Mono_Phago, Phago, Granulo_Phago. Pour cet ensemble de caractères variés, mais biologiquement lié à la phagocytose, nous avons pu identifier 45 sondes dont la variation d'expression semble pouvoir prédire la variation des caractères significativement ($Q_2 = 0.0978$) (Figure E3-39).

Cette approche sera poursuivie pour cribler l'ensemble des paramètres immunitaires.

7. Identification de marqueurs génétiques

a) Intégration avec les résultats eGWAS (Etude 1)

En comparant les sondes dont les niveaux d'expression ont été trouvés associés à des variations de caractères immunitaires à celles trouvées associées à des SNP dans l'étude 1, nous avons mis en évidence une liste globale de 453 SNP-eQTL candidats pour être des marqueurs génétiques de la variation de 43 caractères immunitaires. Ces SNP-eQTL sont associés à 483 sondes, pour 589 associations. Ces sondes représentent 348 gènes. Le détail par caractères est disponible dans le Tableau Supplémentaire 1.

b) Intégration avec les résultats ASE (Etudes 1 et 2)

La liste des SNP-ASE mis en évidence dans les études 1 et 2 et assignés à des gènes qui correspondent à des sondes dont les niveaux d'expression sont associés à des variations de paramètres immunitaires est présentée dans le Tableau Supplémentaire 1. Ces SNP-ASE sont des marqueurs génétiques candidats pour influencer sur la covariation d'expression des gènes et des phénotypes immunitaires associés.

c) Comparaison avec l'approche GWAS

Une analyse GWAS a été faite pour les caractères immunitaires disponibles (données non publiées). Le Tableau Supplémentaire 2 présente les résultats de ces études d'associations. La Figure Supplémentaire 2 présente les Manhattan Plots des analyses correspondantes. Des QTL ont été détectés pour les sept phénotypes descriptifs (CD4-CD8-, CD4+CD8-, cDC2, CRP, EOS, GB, IGA) et pour deux caractères immunitaires



de réponse (IL8_PIL_48h et IL10_LPS_48h). Nous constatons ainsi que l'approche intégrative avec le transcriptome sanguin comme phénotype intermédiaire présenté tout au long de cette thèse permet d'identifier plus de marqueurs génétiques candidats pour plus de caractères immunitaires.

En comparant plus précisément les résultats de ces GWAS avec les résultats de cette thèse, nous avons identifié 16 SNP en commun. Ces SNP sont à la fois associés par GWAS à ces caractères immunitaires et sont associés à l'expression de gènes associés à des caractères immunitaires. Par exemple, les SNP DIAS0004552 et ASGA0021034 ont été associés à la CRP par GWAS. Ils ont été associés à la sonde 07846, elle-même associée au caractère Granulo_Phago. Le SNP MARC0010515 a été associé à la CRP par GWAS. Il a été associé à la sonde A_72_P471568, elle-même associée aux caractères CD4-CD8-, CD4+CD8- et CD4+CD8+. Les SNP MARC0012679 et ALGA0077012 ont été associés au caractère CD4-CD8- par GWAS. Ils ont été associés à la sonde A_72_P471568 annotée pour le gène XIRP2, elle-même associée aux caractères CD4-CD8-, CD4+CD8- et CD4+CD8+.

D. Discussion

Les résultats mettent en évidence trois types d'informations :

1. Une variabilité importante d'une large palette de paramètres immunitaires mesurés à partir d'échantillons sanguins dans des populations porcines sélectionnée de manière intensive sur des paramètres de production ;
2. Des covariations entre paramètres immunitaires qui posent des questions sur la combinatoire des paramètres à mesurer pour qualifier la compétence immunitaire ;
3. La pertinence du transcriptome sanguin comme source de biomarqueurs de la variation des paramètres immunitaires et l'intérêt de coupler des approches eGWAS et GWAS pour identifier des marqueurs génétiques qui influencent sur les niveaux des paramètres immunitaires.

1. Une variabilité des paramètres immunitaires héréditaires, favorable pour leur prise en compte dans des programmes de sélection

L'étude a été conduite sur des porcs Large White intensivement sélectionnés pour des caractères de production (Sanchez *et al.*, 2014), qui incluent des phénotypes liés à l'efficacité alimentaire, la conformation de carcasse, la production de viande et la reproduction. Sans savoir encore comment relier les variations de paramètres immunitaires à une compétence immunitaire effective, nous confirmons dans cette étude que



ces phénotypes immunitaires présentent une variabilité individuelle importante, comme déjà publié précédemment par notre équipe (Flori *et al.*, 2011). Une large gamme de ces paramètres est significativement héritable, mettant en évidence un contrôle génétique potentiellement exploitable en élevage. Il sera essentiel d'évaluer l'impact sur des niveaux de production d'une sélection qui cible l'amélioration de performances immunitaires. En effet, il est couramment admis qu'il existe des compromis fonctionnels entre caractères de production et de santé, qui pourraient complexifier la sélection simultanée sur les deux types de caractères. C'est tout l'enjeu de la multi-performance en élevage dans une perspective agro-écologique (Phocas *et al.*, 2016). Des corrélations positives et négatives ont été identifiées entre paramètres de l'immunité et de croissance (Clapperton *et al.*, 2009 ; Bidanel *et al.*, 2013), sans conclusion nette sur les effets défavorables ou non en sélection. Il est raisonnable de considérer que la santé participe au bien-être animal et donc à l'acceptabilité des systèmes d'élevage par la société, il est fort probable que promouvoir la durabilité aura un coût à court terme, compensé sur le long terme.

2. L'immunophénotypage : une combinatoire complexe de mesures à simplifier ?

La palette de paramètres immunitaires mesurables chez le porc est étendue et couvre des mesures liées à l'immunité innée comme adaptative. La gamme des dosages réalisés correspond pour partie à celles qui peuvent être déployées chez l'homme pour qualifier le niveau basal immunitaire, en tant qu'informations prédictives au cours de la vie (Albers *et al.*, 2005; Calder, 2007; Calder and Kew, 2002). Les projets « Milieu Intérieur » (Thomas *et al.*, 2015) ou ImmVar (De Jager *et al.*, 2015) en sont des illustrations, avec des dosages de paramètres immunitaires à partir d'échantillons de sang stimulé ou non par divers agents dont des pathogènes (Duffy *et al.*, 2014), et la caractérisation de sous-populations cellulaires par cytométrie de flux (Chen *et al.*, 2015; Hasan *et al.*, 2015). L'immunophénotypage à grande échelle mise en place dans des projets ambitieux chez l'homme peut se déployer à la même échelle chez le porc, avec des mesures à la fois nombreuses et fines, y compris sur des sous-populations cellulaires.

La question est maintenant de savoir ce qu'il serait important de phénotyper, en routine et à coût raisonnable, pour qualifier la compétence immunitaire chez le porc. Nos résultats mettent en évidence des corrélations positives et négatives significatives entre paramètres immunitaires, à la fois nombreuses mais plutôt faibles dans l'ensemble, à l'exception notamment des corrélations entre taux de sous-populations cellulaires, ce qui confirme les résultats déjà obtenus (Flori *et al.*, 2011). Ici, nous apportons des informations nouvelles, avec des corrélations positives significatives entre paramètres de phagocytose et entre paramètres qui caractérisent la capacité de production ex vivo de cytokines à différents temps et suite à



diverses stimulations. Ces données suggèrent que la capacité de production des cytokines est répétable pour divers stimuli, et participe ainsi à la qualification de la compétence immunitaire. Nos résultats devraient permettre d'identifier les mesures redondantes et la combinatoire des plus pertinentes. A ce titre, la classification hiérarchique non supervisée réalisée dans cette étude permet de regrouper des mesures et de rechercher des biomarqueurs partagés. Ces résultats sont une base de réflexion pour avancer dans cette direction et identifier, d'une part des mesures à conjuguer sous forme d'un index de sélection et, d'autre part, des biomarqueurs associés. Rappelons ici les travaux pionniers conduits au Canada sur la sélection divergente sur plusieurs générations de porcs avec haut ou bas niveau de réponse immune, sur la base d'un index qui combinait réponse cellulaire et humorale (Mallard *et al.*, 1992), avec de meilleures performances zootechniques pour les animaux de la lignée immune haute (Mallard and Wilkie, 2007). Il existe actuellement un test de compétence immunitaire chez les bovins (HIR pour *High Immune Response*), sous licence d'exploitation, qui se base sur cette nature d'index (Mallard *et al.*, 2015).

3. Vers l'identification de marqueurs génétiques et de biomarqueurs prédictifs des réponses immunes

Nos résultats mettent en évidence des gènes dont la variation d'expression dans le sang est associée à celle de paramètres immunitaires, ces gènes pouvant être des biomarqueurs candidats pour prédire des niveaux immuns, surtout ceux que sont validés par rapport à notre travail précédent (Mach *et al.*, 2013). Certains de ces gènes sont soumis à une régulation génétique de leur expression, comme révélé par les eQTL agissant en local ou à distance et des polymorphismes ASE. Le couplage de ces données avec les GWAS contribuera à prioriser des polymorphismes candidats, cette approche intégrée ayant déjà montré sa puissance, par exemple sur les travaux de Aguirre-Gamboa *et al.* (2016) et Simon *et al.* (2016). Les données générées devraient fournir, d'une part des biomarqueurs sanguins à utiliser comme des phénotypes moléculaires intermédiaires et, d'autre part, des marqueurs génétiques associés aux variations de paramètres immunitaires. Se posera la question de la robustesse des marqueurs prédictifs identifiés, avec en particulier leur portabilité dans d'autres environnements et pour d'autres génétiques que ceux étudiés ici. Des analyses à grande échelle sur des populations variées et dans des environnements différents (élevages commerciaux, statuts sanitaires, pression en pathogènes) seront indispensables pour valider et enrichir les résultats pour identifier des leviers utiles en élevage.



Extrait du Tableau Supplémentaire 1 : Sondes associées aux caractères immunitaires

Sonde_modif	Pheno	AveExpr	F	P.Value	adj.P.Val	Gene	Nombre SNP-eQTL
A_72_P430414	Granulo_Phago	5.3	4.9	0.0002	0.020	TKTL2	5
A_72_P145856	GRA	9.1	4.8	0.0003	0.027		0 4
A_72_P145856	GRA_LYM	9.1	4.6	0.0005	0.038		0 4
A_72_P145856	LYM	9.1	4.57	0.0006	0.046		0 4
A_72_P078761	CD14-CD163-	7.7	5.5	6.8e-05	0.019	APP	3
A_72_P303329	CD14-CD163-	6.9	4.9	0.0002	0.044	CXCL16	3
A_72_P047826	CD14+CD163+	5.5	7.09	3.36e-06	0.003	CMKLR1	3
O13555	CD14+CD163+	7.2	6.1	2.22e-05	0.013	SSH2	3
A_72_P047826	GRA	5.5	5.0	0.0002	0.02	CMKLR1	3
A_72_P047826	GRA_LYM	5.5	4.6	0.0005	0.03	CMKLR1	3
A_72_P150416	Granulo_Phago	10.7	4.5	0.0005	0.039	BCCIP	3
A_72_P191526	IL10_PIL	5.9	8.9	4.4e-07	0.008	USP34	3
A_72_P191526	IL17	5.8	38.6	3.50e-28	2.0e-23	USP34	3
A_72_P315633	IL8_PI	4.2	16.0	1.36e-13	4.0e-09	MUM1L1	3
A_72_P315633	IL8_PIL	4.2	34.2	4.82e-26	1.44e-21	MUM1L1	3
A_72_P133021	MON	11.2	4.5	0.0006	0.043	ZDHHC16	3
A_72_P150416	Mono_Phago	10.7	5.1	0.0001	0.027	BCCIP	3
A_72_P373788	Mono_Phago	8.8	4.5	0.0006	0.046	DHX16	3
A_72_P429579	Mono_Phago	9.76	5.8	4.064e-05	0.015	TXNL4A	3
A_72_P344188	Mono_Phago_MFI	9.2	7.3	2.069e-06	0.0008	PAFAH2	3
A_72_P040566	Phago	6.3	4.5	0.0005	0.02	SSH2	3
A_72_P133021	Phago	11.2	4.2	0.001	0.03	ZDHHC16	3



Extrait du Tableau Supplémentaire 2 : Gènes sous contrôle génétique et associés à des caractères immunitaires

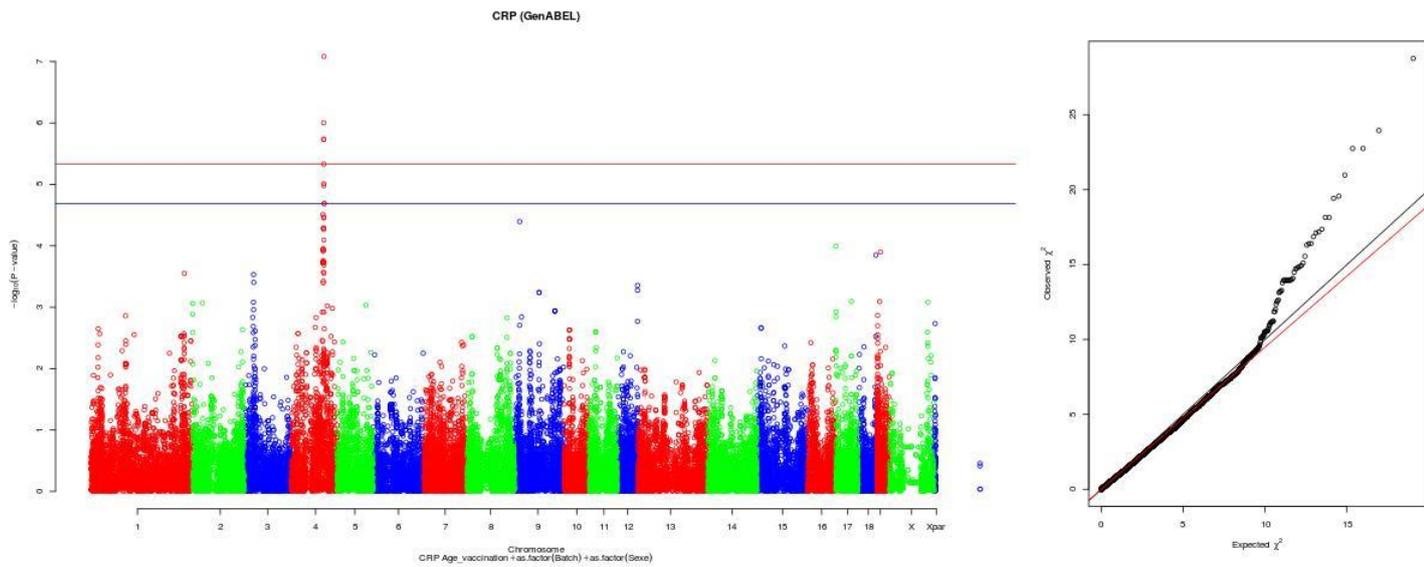
Phentoy pes	Sondes associées avec eQTL (publié)	Genes associés + eQTL	Nombre de gènes associés avec ASE (Etudes 1 et 2)	Gènes associés + ASE
IL2_4h	1	CCL20	0	0
IL1b_4h	3	SLC44A4;ZNF638;TACC3	2	LPL;TACC3
IL8_4h	3	CACUL1;RANBP17;A2M	1	NF1
IL10_4h	0	0	0	0
IL17_4h	1	USP34	1	CHD3
INFg_4h	7	EEF2K;CETN2;TNRC6B;SLC24A1;DEPDC5;STARD5;HINT3	40	AGK;ANKRD44;ATP5A1;B4GALT1;CA5B;CHD2;DCTPP1;DENND2D;DENND4C;FBXO9;FSD1;GABARAPL1;GHITM;GMPR;GNAS;GPI;HINT3;HSP90AB1;IFIT1;IL2RG;IL7R;JAK1;KDM5C;LDHB;MMP9;NTAN1;PRDX2;PSMB7;RPL4;RPL6;RPSA;RRAGA;SETD2;SFSWAP;SLA;SNU13;TMBIM6;TPM4;TRIM26;TUBA1B
TNFa_4h	0	0	0	0
IL1b_LPS_48h	7	SOD3;RBM47;ATAD5;CADM2;MEIS1;EPB41L4A;TCN1;FAM160A2;PSCA;PTCD1	5	HP;SELL;TCIRG1;TEC;UBA5
IL1b_PIL_48h	1	HSDL1;CYP4F11	0	0
IL1b_PIL_48h	2	TRIM15;CYP4F11	1	NAGK
IL8_LPS_48h	8	MIA3;TMEM232;KDM5A;TRIM58;SLC44A4;SAXO1;ADARB2;ZNF638;PLA2G2D	12	ACTR5;ANKRD1;CAPN1;F10;FN1;GPBP1;IMPACT;MIA3;RSF1;SOX13;TRIM58;TTC25



Extrait du tableau Supplémentaire 3 : Résultats des GWAS

SNP	Trait	Chromosome	Position	Strand	A1	A2	N	FDR, Pc1df
MARC0010515	CD4-CD8-	14	40197114	+	G	A	294	8,17E-02
MARC0012679	CD4-CD8-	14	40597533	+	G	A	294	8,17E-02
ALGA0077012	CD4-CD8-	14	40614531	+	A	G	294	8,17E-02
ASGA0103262	CD4-CD8-	6	75138271	+	G	A	294	4,42E-02
ASGA0032106	cDC2	7	29197945	+	G	A	513	1,25E-04
ALGA0039819	cDC2	7	30062235	+	A	G	513	4,41E-02
ALGA0039786	cDC2	7	29771985	+	G	A	513	4,41E-02
ALGA0039770	cDC2	7	29226357	+	A	G	510	7,22E-02
ALGA0026828	CRP	4	101428842	+	G	A	511	3,22E-03
ALGA0026791	CRP	4	100771814	+	A	G	509	1,80E-02
ALGA0026790	CRP	4	100759503	+	A	G	511	1,80E-02
MARC0042655	CRP	4	100787674	+	A	G	511	1,80E-02
ASGA0020972	CRP	4	101026527	+	A	G	510	3,65E-02
ALGA0026769	CRP	4	100620769	+	A	G	511	5,87E-02
H3GA0013493	CRP	4	101009461	+	G	A	506	5,87E-02

Extrait de la Figure Supplémentaire S1 : Manhatten plots des GWAS



Discussion générale

L'objectif global de cette thèse s'inscrit dans un programme de recherche qui vise à proposer des pistes d'action pour prendre en compte la capacité immunitaire dans les schémas de sélection porcine afin d'améliorer la robustesse des animaux en élevage, sans surconsommation d'antibiotiques. Plus précisément, nous avons franchi quelques étapes intermédiaires avec les résultats suivants :

- Confirmation de l'intérêt du sang comme un tissu privilégié pour phénotyper un large éventail de paramètres immunitaires, à même de caractériser la compétence immunitaire individuelle.
- Identification de gènes dont la variation d'expression dans le sang est corrélée à celle de traits immunitaires, et qui peuvent être considérés comme des biomarqueurs candidats pour phénotyper la capacité immunitaire.
- Cartographie du contrôle génétique du transcriptome sanguin et mise en évidence de marqueurs génétiques associés à la variation d'expression de biomarqueurs candidats identifiés.
- Spécificité d'expression de l'immunome des tissus lymphoïdes de l'intestin et de sa muqueuse (GALT), notamment des plaques de Peyer iléales et jéjunales.

Dans cette dernière partie du manuscrit de thèse, nous allons dans un premier temps discuter l'intérêt biologique de nos résultats comme nouvelles ressources disponibles pour les communautés de recherche qui utilisent le porc comme espèce d'élevage ou comme modèle biomedical. Puis nous reviendrons sur les choix méthodologiques et stratégies d'analyses employées. Pour finir, nous proposerons des études complémentaires pour approfondir les résultats et avancer dans l'identification de leviers pour améliorer la durabilité des systèmes d'élevage porcins tout en maintenant leur compétitivité.

A. Les phénotypes immunitaires, le transcriptome du sang et son contrôle génétique comme ressources pour la recherche chez le porc

1. Etablissement de nouvelles ressources chez le porc

L'Etude 1 a permis d'établir la première cartographie du contrôle génétique de la variation d'expression des gènes dans le sang périphérique chez le porc. Cette ressource a été mise à la disposition de la communauté scientifique par le dépôt des données d'expression dans des bases de données publiques, et par la création d'une ressource TrackHub permettant de visualiser les SNP-eQTL et les SNP-ASE dans les



navigateurs de génomes des bases de données Ensembl et UCSC (https://www.trackhubregistry.org/search/view_trackhub/AVszgl3NpL2YcDCmVC8X) (Figure Discussion 1). Chez l'Homme, des consortiums tel que le consortium GTEx (<https://www.gtexportal.org/home/>) ont construit ce type d'outil pour permettre une étude systématique de la variation génétique de la régulation de l'expression des gènes dans des tissus humains de référence. Ce type de ressources chez le porc pourra faciliter l'interprétation des études de GWAS pour différents phénotypes pour lesquels le sang constitue un tissu cible ou de remplacement. Les résultats présentés dans cette thèse ont déjà permis, au sein de notre équipe, d'exploiter cette combinatoire GWAS-eGWAS et d'identifier des gènes et mutations candidates associés à des variations de caractères immunitaires (Maria Ballester, non publié, post-doctorat AgreenSkills+).

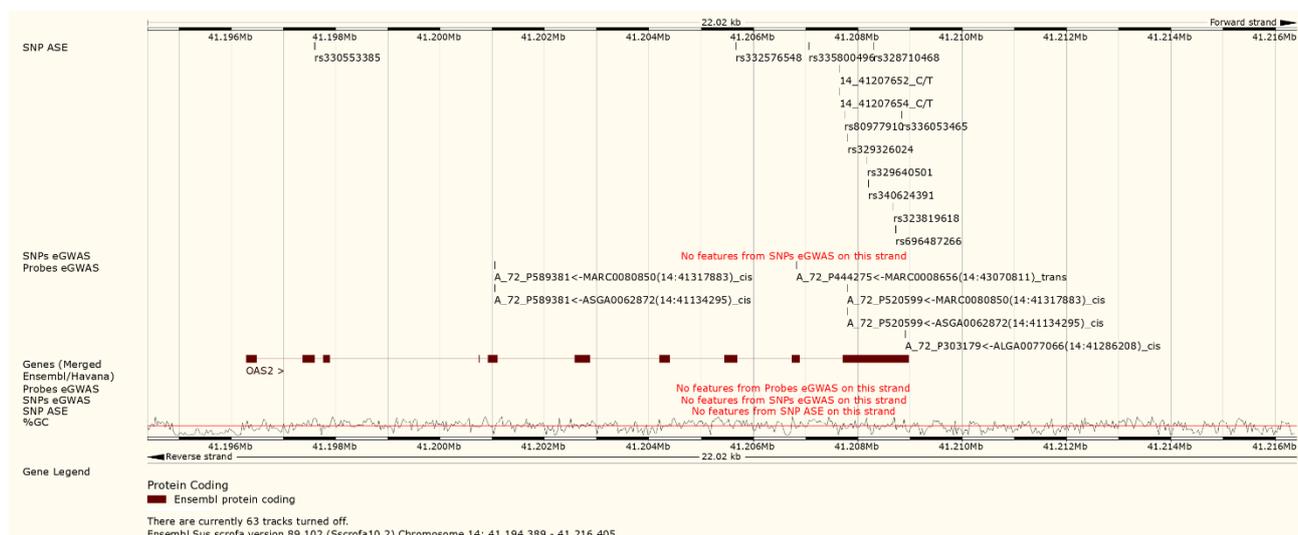


Figure Discussion 1 : Visualisation sur le *Genome Browser* d'Ensembl de la région du gène *OAS2*. Les différents signaux ASE et eQTL sont affichés en haut de l'image grâce aux ressources TrackHub rendu disponible dans l'étude.

L'Etude 2 a permis d'étudier les profils transcriptomiques de tissus lymphoïdes du GALT (ganglion mésentérique, plaques de Peyer iléales et jéjunales) nécessitant l'abattage des animaux (70 jours d'âge), et de les comparer au transcriptome sanguin (à 60 jours d'âge). Jusqu'alors, le profil d'expression génique de ces tissus n'avait été étudié chez le porc que pour de petits nombres de gènes choisis a priori pour leurs fonctions, par exemple Gourbeyre *et al.* (2015). Ici, nous avons proposé une étude du transcriptome complet, incluant l'analyse de la transcription antisens. Cette étude, outre l'analyse fine des différences de profils d'immunome, notamment entre les deux types de plaques de Peyer, a mis en évidence une corrélation



importante entre le transcriptome du sang et celui du GALT. Cela confirme la faisabilité des études d'expression de tissus cibles sans biopsie ou abattage au travers du sang comme tissu de remplacement.

Avec l'étude 3, nous avons analysé 55 caractères immunitaires mesurés directement ou indirectement à partir du sang. Nous avons pu en décrire la variation dans une population sans signe clinique d'infection, reflétant ainsi la variabilité individuelle, à un niveau considéré comme « basal », de ces caractères. Nous avons étudié la co-variation de ces caractères, leur héritabilité et leur sensibilité à des facteurs tels que la bande, le sexe ou encore l'âge à la vaccination. Ces résultats fournissent des données intéressantes pour le choix de caractères immunitaires sanguins sous contrôle génétique, à inclure dans de futures études. Nous avons également cherché à identifier des gènes dont la variation d'expression est associée à la variation des caractères immunitaires sanguins dans notre population d'étude. Ces données constituent une ressource intéressante pour l'identification de biomarqueurs sanguins candidats, et enrichissent nos précédents résultats (Mach *et al.*, 2013). Enfin, par l'intégration des résultats de nos trois études, nous avons pu établir une liste de marqueurs génétiques contrôlant l'expression de gènes associés à des caractères immunitaires, constituant ainsi un ensemble de marqueurs candidats à la sélection sur des critères immunitaires.

2. Intérêt pour les communautés scientifiques en production animale et dans le domaine biomédical

L'ensemble des ressources et résultats présentés dans ce document contribuent à l'acquisition de connaissances sur l'architecture génétique des caractères d'immunité, comme prérequis indispensables pour identifier des leviers d'action en élevage. L'acquisition de ces connaissances a un intérêt au-delà des applications potentielles en élevage. En effet, le porc est également un modèle animal utilisé par la communauté biomédicale dans de nombreuses études fondamentales ou précliniques, car il possède des caractéristiques physiologiques intéressantes et parfois plus pertinentes pour l'homme que le modèle souris (Lunney, 2007) : maturité sexuelle précoce, portées nombreuses, intervalle intergénérationnel relativement court, existence de lignées de mini-porcs, durée de vie plus longue que la souris. De plus, des similitudes avec l'homme facilitent l'extrapolation des études conduites chez le porc : taille des organes, physiologie générale, régime alimentaire et besoins nutritionnels, anatomie, système immunitaire, sensibilité à des infections humaines (Fang *et al.*, 2012).

Le porc a des caractéristiques propres à prendre en compte mais qui peuvent s'avérer être des sujets de recherche d'intérêt, notamment en immunologie comparée (Mair *et al.*, 2014; Rothkötter, 2009). Par exemple, il semble qu'il n'y ait pas de transfert prénatal d'anticorps maternels au fœtus (Salmon, 2011), le



taux de cellules T mémoires CD4+CD8+ est plus élevé chez le porc que chez l'homme, ainsi que le taux de lymphocytes $\gamma\delta$. Le porc se révèle un excellent modèle pour l'étude de pathologies pulmonaires inflammatoires (Maisonasse *et al.*, 2016; Meurens *et al.*, 2012). Le porc est également un modèle bien identifié par les chirurgiens pour élaborer des protocoles de transplantation.

L'utilisation du porc comme modèle pour la recherche biomédicale nécessite des études en profondeur de la biologie porcine et nos travaux permettent de mieux la caractériser. Nos résultats participent donc à une meilleure caractérisation des aspects génétiques, génomiques et immunitaires du porc pour une meilleure exploitation des modèles porcins.

3. Des ressources et des résultats à valider et à approfondir

a) Valider les marqueurs génétiques

Au cours de ce travail de thèse, nous avons établi une cartographie du contrôle génétique du transcriptome grâce à une étude eGWAS et deux études ASE. La combinaison de ces trois analyses permet déjà une validation des résultats. Ainsi, entre les deux analyses ASE, nous retrouvons près de la moitié des signaux en commun. La combinaison des approches ASE et eGWAS a également confirmé, par validation croisée, le contrôle génétique de l'expression de près de 150 gènes. De plus, nous avons validé expérimentalement quelques signaux ASE identifiés lors de l'étude 2, sur les quatre porcs du projet SUS_FLORA, dont certains signaux (gène *OAS2*) avaient déjà été détectés lors de l'étude 1 sur 242 animaux du même projet. Pour valider un SNP-eQTL, c'est-à-dire confirmer l'association entre le SNP marqueur et la variation de l'expression d'un gène, il faudrait, dans l'idéal, sélectionner un groupe d'animaux composés des sous-groupes équilibrés d'animaux présentant les génotypes hétérozygotes (Aa) et homozygotes (AA et aa) pour le SNP en question. On pourrait mesurer le niveau d'expression du gène associé par RT-qPCR et effectuer un test d'association génétique suite au génotypage du SNP dans une population indépendante. Si la différence de moyenne d'expression est conservée, le SNP marqueur de la variation d'expression est validé.

Parmi ces marqueurs génétiques de l'expression des gènes dans le sang (SNP-eQTL ou SNP-ASE), certains ont pu être sélectionnés comme marqueurs génétiques candidats associés à des caractères immunitaires lors de la troisième étude. Il serait intéressant de valider quelques uns de ces candidats. Il faudrait mesurer le caractère immunitaire considéré et comparer la valeur du caractère pour chaque génotype dans une population indépendante puis poursuivre par des études fonctionnelles pour identifier le mécanisme d'action de la mutation si elle est causale.



b) Identifier les mutations causales

Comme décrit précédemment, les SNP que nous avons identifiés sont des marqueurs associés à la variation d'expression des gènes mais ne sont probablement pas causaux. Cependant, pour des SNP-marqueurs validés, il est envisageable de développer des approches de cartographie fine et de chercher à identifier le variant causal ou QTN (*Quantitative Trait Nucleotide*). Dans le cas d'une eGWAS pour un signal *cis* ou d'une approche ASE, et contrairement aux GWAS, le gène candidat impliqué dans le mécanisme de régulation est déjà identifié puisqu'il s'agit du caractère étudié. Il faudra cependant (i) amplifier et séquencer les régions régulatrices et codantes du gène ; (ii) détecter les SNP dans ces régions ; (iii) refaire une analyse d'association afin d'identifier le SNP le plus associé qui pourrait être causal.

Des études fonctionnelles seront ensuite nécessaires pour confirmer la causalité et en identifier le mécanisme régulateur. De nombreuses méthodes existent et le choix de l'une d'elles dépend de la position de la mutation causale candidate. Par exemple, si la mutation est dans le promoteur du gène, il est possible de faire d'abord une analyse *in silico* pour prédire des sites de fixation de facteurs de transcription. Il est également possible d'utiliser des approches de transfection pour étudier l'impact de la mutation sur l'activité du promoteur. Si la mutation est dans une région codante, il faudra analyser son effet sur l'activité de la protéine. Finalement, si elle se trouve dans la région 3'UTR, il faudra s'intéresser à la stabilité de l'ARNm et étudier si des sites de fixation de miARN sont modifiés selon les génotypes au site de la mutation candidate.

c) Valider les biomarqueurs sanguins

Mach *et al.* (2013) ont déjà identifié des corrélations entre la variation du transcriptome sanguin et celle de caractères immunitaires mesurés dans le sang et identifié des biomarqueurs candidats pour qualifier les taux de lymphocytes CD4-CD8+, la capacité phagocytaire, la production d'IL-2 après stimulation *ex vivo* du sang total, sur un groupe d'animaux du projet IMMOPIG. Nous avons pu valider notre approche en confirmant l'intérêt de gènes tels que GATM, SCARB1, DDIT4, GZMB et CXCR1 comme des biomarqueurs candidats pour des caractères immunitaires. Nous avons même pu confirmer l'association entre CXCR1 et le caractère CD4-CD8+.

D'autre part, pour une validation plus large, une première approche rapide pourrait être menée sur les animaux SUS_FLORA non inclus dans cette analyse, qui constituerait ainsi une population additionnelle de validation de plus de 300 animaux, pour lesquels la plupart des différents caractères immunitaires sanguins ont déjà été phénotypés. Il s'agirait alors de mesurer spécifiquement la valeur d'expression des biomarqueurs



candidats identifiés par RT-qPCR par exemple, dans les échantillons de sang conservés à -80°C dans des tubes PAXgene, et de valider la capacité prédictive des niveaux d'expression génique par rapport aux phénotypes immunitaires (Mach *et al.*, 2013).

Ensuite, il sera indispensable de tester ces biomarqueurs candidats dans d'autres populations en races pures comme en croisements, dans divers environnements, afin d'estimer la robustesse des biomarqueurs et leur potentiel d'application. Les animaux utilisés dans nos études ont été produits et élevés dans des stations expérimentales de l'INRA, à haut statut sanitaire. Il faudrait évaluer l'impact du statut sanitaire sur les approches que nous avons menées, en utilisant des animaux élevés en conditions de production commerciale. Il faudrait ainsi élargir leur protocole de suivi à des mesures transcriptomiques à haut débit et tester s'il est possible d'identifier les mêmes corrélations et associations entre les polymorphismes, le transcriptome et les phénotypes immunitaires. L'organisation pyramidale de la production porcine donne accès à des élevages de sélection à très haut statut sanitaire avec en général aucun cas de maladie depuis plusieurs années. Les élevages de multiplication et de production ont un statut sanitaire plus faible et pourraient fournir des données diversifiées intéressantes. Il serait également pertinent de phénotyper des animaux dans des élevages de plein-air et des élevages labélisés agriculture biologique.

B. Choix méthodologiques et limites des approches

1. Etude du transcriptome du sang

a) Spécificités du transcriptome du sang

Le transcriptome du sang est riche en ARN issus de gènes majeurs sanguins : la globine et la ferritine. Ainsi dans le protocole de RNA-Seq, ces gènes séquestrent un pourcentage important de lectures, ce qui peut limiter l'estimation de l'expression de transcrits peu abondants si la profondeur de séquençage est faible. Par exemple, dans notre étude 2, les gènes de l'hémoglobine (HBA et HBB) ont globalisé 70 % des lectures.

Des protocoles de déplétion de la globine sont proposés afin de limiter ce biais (Choi *et al.*, 2014), qui n'impactent pas globalement la qualité des ARN, malgré une diminution du RIN pour certains échantillons. La déplétion de globine peut réduire significativement la quantité d'ARN isolés sans toutefois avoir d'effet problématique sur la construction des libraires d'ADN complémentaires à partir des ARN extraits (Shin *et al.*, 2014). Le nombre total de lectures et de lectures alignées comme le pourcentage de lectures filtrées ne semble pas non plus affecté. La déplétion de globine permet ainsi la détection d'un plus grand nombre de transcrits avec plus d'un millier de gènes surexprimés après la déplétion (Shin *et al.*, 2014). Il a cependant été



mis en évidence qu'un groupe de transcrits était déplété en même temps que ceux de la globine (Choi *et al.*, 2014; Shin *et al.*, 2014), la plupart de ces transcrits étant non codants avec une forte variabilité d'expression entre les individus. Cependant, le protocole de déplétion de globine coûte cher. Il est admis qu'une augmentation de la profondeur de couverture permet de compenser correctement l'absence de déplétion de globine (Shin *et al.*, 2014).

Le sang est un tissu composite, avec de multiples populations cellulaires en proportions variables selon de nombreux facteurs. Ainsi, l'étude de transcriptome du sang complet est le reflet du transcriptome de ces sous-populations cellulaires et impose de travailler sur le profil d'expression, non pas d'un seul type cellulaire, mais sur un compromis de niveau d'expression de différents types cellulaires, qui va pouvoir changer en fonction des proportions cellulaires mais également de l'état des cellules (activées, dormantes, sénescentes, naïves, matures, etc.). Certaines études du transcriptome du sang utilisent des données de comptage cellulaire pour corriger les niveaux d'expression. D'autres études se concentrent sur le transcriptome de populations cellulaires bien individualisées après tri. Dans nos études, nous avons fait le choix de ne pas corriger par la composition cellulaire, parce que nos animaux proviennent tous d'un système d'élevage à environnement et alimentation contrôlés, et que les données de numération de formule sanguine n'ont pas révélé d'infection majeure chez nos individus. Nous avons donc considéré le transcriptome complet du sang périphérique pour identifier des biomarqueurs sanguins, sans corriger par les taux de sous-populations cellulaires.

b) Régulation

Le transcriptome est un phénotype intermédiaire soumis à de nombreux mécanismes de régulation et son contenu ne se résume pas à la simple transcription des gènes qui codent des protéines. Les ARN non codants, longs et courts représentent une part significative du transcriptome et ils sont à considérer pour la recherche de biomarqueurs (Grubbs *et al.*, 2016). Par exemple, Xia *et al.* (2017), ont proposé d'utiliser pour l'étude des PBMC les ARN longs non codants (ARNlnc) comme un phénotype intermédiaire complémentaire à l'expression des gènes pour faciliter encore l'étude des mécanismes fonctionnels sous-jacents, en menant tout d'abord une eGWAS sur les ARNlnc puis une analyse de réseaux pour l'intégration et enfin un test d'inférence causale pour identifier les chaînes de régulation au sein d'un trio ARNlnc-ARNm-SNP. Dans l'étude 2, nous n'avons pas inclus l'analyse systématique des ARN longs non codants ou des miARN mais avons pris en compte la transcription antisens.



2. Identification de marqueurs génétiques de l'immunité en deux étapes

a) Approches eGWAS

- Définition d'un eQTL : *cis* vs. *trans*

L'approche eGWAS permet de mettre en évidence des régions génomiques associées à la variation de l'expression d'un gène. Chaque eQTL détecté est délimité sur le génome par un groupe de SNP localisés dans une même région, associés à la variation d'expression d'un gène (ou transcrit) de façon significative. L'identification d'un eQTL correspond donc à une région du génome bornée par des SNP associés les plus significatifs. Souvent, les eGWAS mettent en avant des eQTL représentés par un unique SNP, nommé selon les études tagging-SNP (Benton *et al.*, 2013) ou eSNP (Mehta *et al.*, 2013). Ce SNP représentatif peut être le plus significatif de la région (Benton *et al.*, 2013). Dans notre étude, nous avons choisi de représenter chaque eQTL par le SNP le plus significatif et représentatif de l'effet, et de le nommer SNP-eQTL (eQTL-SNP en anglais). La méthode, inspirée de l'approche utilisée par Bryois *et al.* (2014) et Fairfax *et al.* (2014), a consisté à inclure le SNP le plus significatif de chaque pic d'association dans le modèle statistique et à ré-évaluer l'association pour chaque SNP du eQTL par régression linéaire : si l'association n'est plus détectée, alors le SNP est éliminé car il représente le même effet eQTL que le SNP le plus significatif. Dans le cas contraire, il est conservé comme un eQTL-SNP représentant l'effet d'un autre eQTL.

Comme décrit dans l'introduction, l'influence génétique peut être *cis*, c'est-à-dire avec un effet allélique direct entre le polymorphisme et le gène, ou *trans* si l'effet est indirect, en agissant sur un élément régulateur d'une façon allèle spécifique qui va réguler l'expression du gène associé. Il peut être proche ou éloigné du gène affecté, ou sur un autre chromosome. L'approche eGWAS permet de détecter les eQTL contenant des variants causaux avec un effet *cis* ou un effet *trans*, mais pas de les différencier clairement. Plusieurs études ont donc fait le choix de fixer arbitrairement une limite de distance pour distinguer les eQTL *cis* et *trans* : effet *cis* à moins de 100kb (Zou *et al.*, 2012), 250 kb (Fehrmann *et al.*, 2011), 500kb (Mehta *et al.*, 2013), 1Mb (Sasayama *et al.*, 2013), 2Mb (Bryois *et al.*, 2014; Wang *et al.*, 2014) ou même 10Mb (Kelly *et al.*, 2014) du gène. Cependant, cela peut entraîner une mauvaise classification des eQTL *cis* situés à distance du gène ou de *trans* très proches.

De plus, la plupart des eGWAS ont observé un plus fort effet des eQTL *cis* que des *trans*, reflétant en partie une puissance de détection plus faible pour ces derniers. Ainsi, afin d'augmenter la puissance de



détection des effets *trans*, des études ont mené deux eGWAS en séparant les SNP proches du gène des SNP à distance, obtenant dans le premier cas des eQTL *cis* et dans le deuxième cas, des eQTL *trans* (Bryois *et al.*, 2014; Fehrmann *et al.*, 2011; Landmark-Høyvik *et al.*, 2013; Mehta *et al.*, 2013; Wang *et al.*, 2014).

Dans notre Etude 1, nous avons fait le choix de ne mener qu'une analyse globale et de caractériser les eQTL en deux autres catégories comme suggéré par plusieurs publications (Gilad *et al.*, 2008; Kelly *et al.*, 2014) : local si l'eQTL-SNP se situe à moins de 1Mb de la position de la première base alignée de la sonde représentant le gène et distal s'il est à plus de 1Mb ou sur un autre chromosome (Figure Discussion 2). Les eQTL qualifiés de locaux seront plus probablement des eQTL à effets *cis* et les distaux des *trans*. Toutefois, la validation de l'effet *cis* ou *trans* demandera des études complémentaires pour valider le mécanisme moléculaire de contrôle de l'expression du gène par le polymorphisme.

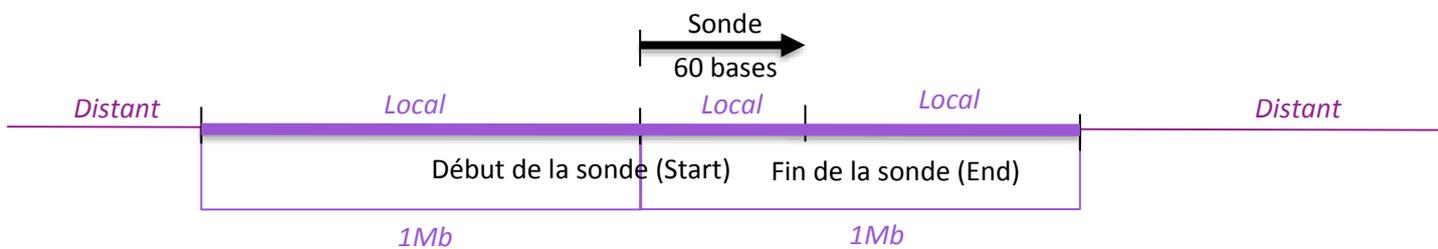


Figure Discussion 2 : Représentation de la répartition des eQTL dans les catégories « local » ou « distal » en fonction de leur position par rapport à la position de la sonde alignée sur le génome

- eQTL à effets multiples

Les eGWAS permettent d'identifier des eQTL associés à l'expression simultanée de différents transcrits ou gènes. Ils sont qualifiés de régulateurs majeurs (*master regulator* en anglais) (Fehrmann *et al.*, 2011). Une façon simple de les identifier est de représenter les associations SNP-gène grâce à un logiciel de visualisation de réseaux. Ainsi, chaque SNP-eQTL est relié à un gène. Notre approche reposant sur la représentation d'un eQTL par le SNP le plus représentatif a pu limiter l'identification de marqueurs génétiques dits majeurs, c'est-à-dire associés à la régulation de multiples gènes. En effet, la significativité d'un SNP-eQTL dépend en partie du déséquilibre de liaison avec le variant causal et donc de sa distance avec lui. Plus il en est éloigné, plus le déséquilibre de liaison sera faible et plus l'association pour le marqueur sera difficile à détecter, et donc moins significative. Ainsi, pour deux gènes biologiquement influencés par la même région génomique voire le même variant causal, il est possible que le SNP-eQTL représentant pourtant le même eQTL ne soit pas le même du fait de la distance au variant causal. Pour être plus exhaustif, nous aurions pu comparer les régions



eQTL associées à différents gènes par leur position sur le génome et considérer deux eQTL chevauchant comme le même, même si le SNP-eQTL n'est pas le même. Nous aurions pu également tester le déséquilibre de liaison entre les SNP-eQTL et considérer les SNP-eQTL en fort déséquilibre de liaison comme un seul eQTL.

b) Approches ASE dans des données RNA-Seq

L'approche ASE à partir des données de RNA-Seq permet de détecter des marqueurs génétiques représentatifs d'un effet *cis*. Comme dans le cas des eGWAS, les SNP avec une expression allèle spécifique sont certainement en déséquilibre de liaison avec le variant causal responsable du contrôle génétique *cis* de l'expression du gène. Les SNP ont été détectés directement dans les données de RNA-Seq et la différence de proportion de chaque allèle a été testée par une approche binomiale.

Dans l'étude 1, nous avons fait le choix de ne sélectionner que les SNP identifiés comme ASE dans au moins un tiers des animaux hétérozygotes, afin de fournir des résultats représentatifs de la population et non d'individus isolés. Dans le cadre de la deuxième étude, nous avons travaillé sur quatre animaux avec quatre tissus analysés par RNA-seq. Nous avons considéré un ASE comme spécifique d'un tissu s'il n'était pas identifié dans un autre tissu, même d'un autre animal. Une autre approche aurait pu être envisagée, comme celle de travailler à l'échelle de chaque animal et de considérer un ASE spécifique d'un tissu s'il n'était pas détecté dans un autre tissu du même animal. De cette façon, la variabilité intra-individuelle aurait été prise en compte. Cependant, la profondeur de lecture des gènes pour chaque tissu d'un même individu n'est pas la même et cela peut inclure un biais dans la définition d'un ASE spécifique au sein d'un seul individu. Edsgård *et al.* (2016) ont récemment proposé une méthode afin de pallier ce type de problème et de pouvoir détecter les ASE spécifiques à la fois d'un individu et de tissus (GeneiASE). Ils ont ainsi proposé un catalogue des ASE statiques (et donc spécifiques d'un individu quel que soit le tissu ou la condition) et conditionnels (différents selon le tissu ou la condition) pour étudier des leucocytes avec ou sans traitement au LPS. Il sera intéressant de tester cette approche sur nos données.

La détection d'expression de gènes spécifiques d'allèle à partir de données de RNA-Seq permet d'identifier des phénomènes d'empreinte parentale liés à un contrôle épigénétique (DeVeale *et al.*, 2012). Notre analyse ne permet pas de différencier les ASE soumises à empreinte ou non faute d'avoir accès au génotype et/ou à la séquence complète du génome des animaux étudiés et de leurs parents. Cela nous aurait permis d'observer si l'allèle surexprimé provenait toujours du même parent.

Une importante limite de l'approche ASE est la capacité de ne détecter des ASE que pour les SNP présents dans la population sous forme hétérozygote et avec une fréquence suffisante. Dans notre première



étude, nous avons des données RNA-Seq pour 38 animaux. Sans accès ni à du génotypage haute densité ni au séquençage génome entier, nous avons identifié les SNP présents dans cette population uniquement à partir des séquences issues du RNA-Seq. Cela implique de nombreux faux positifs car les transcrits sont soumis à des mécanismes d'épissage et des modifications post-transcriptionnelles. De plus, les SNP entraînant l'inhibition complète ou presque de la transcription d'un allèle ne seront pas détectés alors qu'ils correspondent à des SNP-ASE. Si nous avions eu accès au génome entier de chaque animal, nous aurions pu valider les SNP avec des données de séquençage de l'ADN du génome. Cependant, afin de limiter cet effet, nous avons établi des filtres stricts et comparé les SNP détectés avec ceux référencés dans la base de données dbSNP (Sherry *et al.*, 2001). Cet effet était moins important dans la seconde étude car, malgré seulement quatre animaux et toujours que du RNA-Seq, nous avons des données pour quatre tissus. Ainsi, grâce notamment à l'outil QuASAR (Harvey *et al.*, 2015), nous avons pu génotyper avec plus de précision les animaux pour une liste de SNP détectés toujours dans des données RNA-Seq, en combinant les informations d'expression allélique des quatre tissus.

c) Intégration des données

- Complémentarité entre eGWAS et GWAS

La recherche de marqueurs prédictifs des caractères immunitaires sanguins a été réalisée dans ce projet en deux étapes, en utilisant le transcriptome sanguin comme phénotype intermédiaire : (i) eGWAS pour identifier des marqueurs génétiques de l'expression des gènes ; (ii) covariation des gènes exprimés dans le sang et des caractères immunitaires sanguins. Cela a permis d'intégrer trois niveaux biologiques : génotypes, transcriptome et phénotypes immunitaires.

L'approche directe par GWAS, réalisée parallèlement dans notre équipe mais en dehors des objectifs de cette thèse, s'est révélée insuffisante pour de nombreux caractères, ce qui peut s'expliquer par le fait que le processus qui relie le polymorphisme au caractère final mesuré passe par de nombreuses étapes qui peuvent masquer l'effet du polymorphisme. Par exemple, le consortium GTEx (GTEx Consortium, 2013, 2015; GTEx Consortium *et al.*, 2017) a mis en évidence que la plupart des variants identifiés en GWAS ne sont pas annotés au plan fonctionnels, 90 % d'entre eux étant localisés hors des régions codantes. Ainsi, la cartographie d'eQTL est une approche puissante complémentaire des GWAS pour élucider les liens entre génotypes et phénotypes cibles.

Ce type d'approche a déjà donné des résultats pour des caractères immunitaires. Par exemple, la complémentarité entre des résultats de GWAS et de eGWAS a permis à Aquirre-Gamboa *et al.* (Aquirre-



Gamboa *et al.*, 2016) de décrire des loci associés au nombre de lymphocytes, et Simon *et al.* (Simon *et al.*, 2016) ont établi une association entre le gène *MFN2* et le nombre de plaquettes.

- Méthodologie pour l'identification de signatures de biomarqueurs

Dans l'étude 3 de ce travail de thèse, nous avons identifié des signatures de biomarqueurs à partir du transcriptome sanguin, candidats pour prédire la variation d'expression de caractères immunitaires mesurés à partir du sang. Pour cela, nous avons utilisé une approche sPLS (Lê Cao *et al.*, 2009), nous permettant de choisir le nombre de gènes formant cette signature. Ainsi, toujours dans le but de participer à l'identification d'outils pour sélectionner les porcs sur des critères immunitaires, nous avons décidé de fixer ce nombre de sondes/gènes à 45. En effet, plusieurs technologies existent actuellement pour mesurer précisément l'expression d'un petit groupe ciblé de gènes. L'une d'elle est le système à haut débit de Fluidigm (48.48 Dynamic Array™ IFC ; <https://www.fluidigm.com/products/biomark-hd-system>) qui permet la mesure simultanée de 48 transcrits ou gènes. En prévoyant l'inclusion de trois gènes de référence dans la puce comme contrôle qualité et pour la normalisation des données, nous avons fixé le nombre de gènes pour les analyses sPLS à 45. Dans l'idéal, il serait intéressant de mesurer systématiquement l'expression de tous les gènes mais cela serait encore trop coûteux et demanderait des traitements bioinformatiques assez peu conciliables avec des systèmes de routine.

C. Perspectives en élevage

1. Confirmer l'utilisation de caractères immunitaires sanguins pour qualifier la compétence immunitaire individuelle

Nous avons basé nos études sur des caractères immunitaires mesurés dans le sang chez des animaux sains, avec une approche appelée du « caractère marqueur » par Visscher *et al.* (2002). Il faudra par la suite pouvoir définir dans quelle mesure ces traits immunitaires sont représentatifs des capacités immunitaires réelles des animaux. En effet, il est important de comprendre comment chaque caractère immunitaire influence sur les capacités immunitaires individuelles et prédisent les réponses à des pathogènes ou des stress abiotiques. Comme discuté par Calder (2007) pour l'Homme, il est connu qu'une diminution des fonctions du système immunitaire entraîne une augmentation de la sensibilité aux maladies, de la morbidité et de la mortalité. Cependant, cette immunodépression est un cas extrême et il est moins évident de savoir quel impact peut avoir une légère diminution ou augmentation d'un indicateur des fonctions immunitaires. Il peut



en effet n'y avoir aucun impact sur la résistance globale aux maladies, notamment grâce à la redondance du système immunitaire qui pourrait mettre en place des mécanismes de compensation.

De hauts niveaux de réponses immunitaires ne garantissent pas toujours une meilleure résistance face à des pathogènes variés. La surexpression de réponses immunitaires peut même être contre-productive, pouvant entraîner le développement de maladies auto-immunes notamment. Chez le porc, la mortalité en cas de pneumonie grippale est due à une tempête cytokinique déclenchée en réaction à l'infection par le virus, qui entraîne une destruction des tissus pulmonaires.

Ainsi, chez le porc, comment déterminer le niveau optimal de chaque paramètre immunitaire basal pour optimiser la robustesse des animaux face à un challenge infectieux ? Par exemple est-il préférable d'être en capacité de produire plus ou moins d'IL-2 après stimulation *in vitro* ? Est-il préférable de sélectionner selon la capacité de phagocytose ?

Idéalement, il nous faudrait confronter les variabilités individuelles des paramètres immunitaires mesurés aux réponses suite à des challenges infectieux, des vaccinations ou des stress. Les animaux testés pourraient être soit des populations sélectionnées sur plusieurs générations pour des index de paramètres immunitaires et soumises à des infections pas différents pathogènes en conditions expérimentales, soit des populations commerciales ou d'élevages conventionnels, soumises aux pressions en pathogènes de leur environnement. Dans tous les cas, il faudra étudier la vitesse de mise en place des réactions immunitaires, la résistance, la robustesse et la résilience ou capacité de guérison. Mais cette approche est longue et laborieuse et surtout peu soucieuse du bien-être des animaux. Dans un souci d'éthique en expérimentation animale, il sera essentiel de développer à grande échelle des approches *ex vivo* pour tester la résistance aux pathogènes et la protection des animaux suite à la vaccination.

Au sein de l'équipe GIS, nous sommes impliqués dans divers projets pour avancer dans ce lien entre paramètres immunitaires au niveau basal et compétence immunitaire effective. Nous collaborons avec des équipes canadiennes et américaines sur un projet (Phenomix) qui propose de phénotyper régulièrement des animaux laissés libres dans un environnement ouvert et en mélangeant des générations différentes, permettant ainsi la propagation naturelle de maladies au sein de l'élevage. Le suivi des animaux va permettre d'identifier les animaux les plus robustes face aux pathogènes qui apparaîtront au sein de l'élevage et l'accès à des données précoces de caractères immunitaires, de génotype, de transcriptome permettra d'identifier des marqueurs prédictifs de résilience. Dans le cadre du projet H2020 SAPHIR, nous analysons le transcriptome sanguin (par RNA-Seq et miRNA-Seq) avant vaccination contre *Mycoplasma hyopneumoniae* ou contre le virus de la grippe porcine, afin d'identifier des marqueurs génétiques et des biomarqueurs sanguins prédictifs de la variabilité des réponses anticorps à la vaccination.



Un consortium français, le programme PigletBiota (ANR, 2015-2018), est piloté par notre équipe, impliquant un partenariat INRA avec BIOPORC et des firmes en nutrition animale (Techna France Nutrition, Deltavit, Mixscience, Invivo et Lallemand). Ce projet a pour but l'étude conjointe du microbiote intestinal, du transcriptome sanguin et du génotype des animaux pour qualifier la robustesse au sevrage de porcs Large White et croisés Large White. Les données générées constitueront également une source de validations supplémentaires pour les résultats de cette thèse.

Finalement, un dernier projet en cours est une collaboration avec des équipes italiennes (Université de Bologne et IZSUM) dans le cadre de la thèse en co-tutelle de F.R Massacci. Cette approche consiste à sélectionner des animaux à partir d'une liste de marqueurs génétiques et de soumettre ces animaux à des pathogènes afin d'étudier leur réaction. Donc, les animaux générés représenteront des opportunités de validation des résultats de cette thèse par rapport à leur capacité à prédire non pas la réponse immunitaire mais la robustesse face à des pathogènes.

2. D'autres sources de biomarqueurs non invasives : le microbiote

Le microbiote intestinal est la population de microorganismes vivant dans le tube digestif : bactéries, virus, champignons, etc. Chez l'homme, il représente 10 000 milliards de cellules, soit 10 fois plus de cellules que dans le corps humain. Il a été mis en évidence une coopération entre microorganismes et hôtes avec une participation des microorganismes à des fonctions métaboliques et de défense. L'analyse du microbiote intestinal est un domaine en pleine expansion pour étudier finement la construction des phénotypes et identifier comment sa maîtrise pourrait améliorer des caractères en élevage (Calenge *et al.*, 2014).

Le microbiote intestinal du porc peut potentiellement influencer l'efficacité alimentaire, la capacité immune, la résistance aux maladies, le bien-être. Le catalogue des gènes du microbiome intestinal chez le porc a été récemment publié avec une contribution de notre équipe (Xiao *et al.*, 2016). Cette étude a mis en évidence que les facteurs de l'hôte influent vraisemblablement sur la composition du microbiome intestinal, tels que la génétique, l'âge ou le sexe. Lors du programme SUS_FLORA, des prélèvements de fèces non invasifs ont été réalisés pour caractériser l'abondance relative des genres bactériens par séquençage d'un gène codant pour l'ARNr16S, une sous unité ribosomale (Mach *et al.*, 2015; Ramayo-Caldas *et al.*, 2016). Parmi les résultats obtenus, deux entérotypes ont été identifiés, dominés par les genres *Prevotella* et *Mitsuokella* ou *Ruminococcus* et *Treponema*. L'entérotype dominé par *Prevotella* pourrait être plus favorable à la croissance et à la production d'IgA lumineales (Mach *et al.*, 2015; Ramayo-Caldas *et al.*, 2016). Les deux



entérotypes ont été confirmés dans une autre étude (Yang *et al.*, 2017), suggérant que des biomarqueurs microbiens pourraient également être pris en compte pour qualifier des phénotypes à améliorer en élevage, en particulier la santé.

3. Sélectionner les animaux sur leur compétence immunitaire

Dans ce projet de thèse, nous avons travaillé sur des porcs Large-White Français issus des schémas de sélection porcine classiques français. Ils sont issus d'une lignée pure soumise à une forte pression de sélection ou, depuis 1995, les lignées mâles et femelles ont été sélectionnées selon différents critères (voir Introduction). Parfois, la sélection sur un critère particulier entraîne une pression de sélection sur un autre phénotype, par le jeu des allèles et des effets pléiotropiques. Par exemple, le locus dit Halothane contient une mutation dominante associée à une diminution du lard dans la viande (Fujii *et al.*, 1991). Ainsi, sur des critères de production, elle est intéressante. Cependant, elle est également associée de façon récessive à une hyperthermie maligne déclenchée par le stress ou lors d'une anesthésie par exposition au gaz halothane. Ainsi, la sélection positive de cet allèle récessif doit être maîtrisée pour éviter l'existence en production d'animaux homozygotes récessifs. Cet exemple nous incitera à vérifier que les marqueurs de caractères immunitaires à tester ne seraient pas associés à d'autres phénotypes non désirés dans la production.

Il existe deux stratégies de base pour l'amélioration de la résistance aux maladies : directe et indirecte. La sélection directe a pour but d'améliorer la résistance pour une maladie spécifique pour en réduire la sévérité. La sélection indirecte vise la sélection de phénotypes candidats comme des traits immunitaires. Ce projet de thèse s'inscrit dans la sélection indirecte. Les deux méthodes ont leurs avantages et leurs inconvénients. Ainsi, dans la sélection directe il est souvent difficile d'estimer l'état de la maladie dans une population d'élevage. La sélection indirecte, elle, nécessite de tester la corrélation des phénotypes sélectionnés avec la résistance actuelle à certaines maladies car elle n'est pas basée sur des symptômes infectieux.

La sélection sur des critères immunitaires pose la question de la balance énergétique et du compromis entre fonctions vitales: croissance, digestion, résistance aux pathogènes, reproduction. Il sera donc essentiel de s'assurer que les animaux sélectionnés pour leurs capacités immunitaires ne verront pas leurs autres capacités, par ailleurs essentielles pour la production, impactées négativement. Plusieurs études de sélection sur différents critères immunitaires sur au moins cinq générations de porcs ont déjà pu observer



une amélioration des critères immunitaires sans effet sur des critères de production (Katayama *et al.*, 2005; Mallard and Wilkie, 2007). Des corrélations positives et négatives ont été mesurées chez le porc entre paramètres immunitaires et de croissance (Clapperton *et al.*, 2009 ; Bidanel *et al.*, 2013).

4. *Influence de l'environnement*

Les études génétiques conduites pendant ce projet de thèse se sont appuyées sur des populations à environnement maîtrisé et constant. Pourtant, les phénotypes immunitaires sont le résultat d'interactions fortes entre les facteurs environnementaux et génétiques. Comprendre les processus à l'échelle de l'individu amenant au phénotype mesuré est important non seulement pour comprendre comment la sélection a agi mais aussi pour avoir la possibilité de prédire comment les phénotypes pourraient évoluer face à une modification de l'environnement (Reines *et al.*, 2016). Pour les animaux d'élevage tels que les porcs élevés en milieu contrôlé de production intensive, cet effet d'interaction doit être faible. Mais dans des approches impliquant plusieurs générations, il doit y avoir un effet « année-saison » sur les caractères immunitaires (pour l'Homme : Aguirre-Gamboa *et al.*, 2016). Les approches mêlant des animaux de plein air et/ou des animaux de différentes fermes doivent également en tenir compte. En effet, Clapperton *et al.* (2009) ont pu identifier une différence d'héritabilité pour certains caractères immunitaires en fonction du statut sanitaire de l'élevage porcin. Baud *et al.* (2017) ont également décrit un effet de la variation génétique de l'environnement social sur l'état de santé individuel.

Le système immunitaire est également très influencé par l'âge (Pawelec *et al.*, 1998). Pour l'Homme, Brodin *et al.* (2015) dans une étude des caractères immunitaires chez des paires de jumeaux, au fond génétique identiques ont observé une augmentation de la variabilité de ces caractères avec l'âge. Aguirre-Gamboa *et al.* (2016) ont montré par exemple des changements liés à l'âge dans le système immunitaire humain comme la diminution des cellules B naïves et cellules T, avec une augmentation concomitante des cellules B et T mémoires. Il faudrait donc réfléchir à un temps adéquat pour la recherche de biomarqueurs. Les mêmes auteurs ont également confirmé que le genre a une influence sur le système immunitaire : taux de cellules immunitaires et d'immunoglobulines plus importants chez les femmes, sauf pour les cellules T effectrices et mémoires. Bien que les animaux de production en engraissement sont sacrifiés relativement jeunes, il sera intéressant de déterminer si les biomarqueurs identifiés dans cette thèse sont prédictifs de l'immunité et de la santé d'animaux plus âgés comme les truies reproductrices en élevage.



5. Vers une agriculture de précision qui intègre de multiples mesures

La modification des comportements (réduction des déplacements, posture corporelle, etc) est un signe précoce d'altération de la santé à ne pas négliger, qui requiert des équipements spécifiques de collecte et d'analyse. Les approches traditionnelles du phénotypage comportemental (stratégies d'adaptation, réactivité émotionnelle, tendances agressives, comportements spécifiques) sont encore essentiellement basées sur des observations directes de l'éleveur. Néanmoins, des enregistrements vidéo des animaux, avec une notation manuelle des comportements mais des logiciels d'analyse comportementale à haut débit, se développent considérablement pour permettre dans le futur une prise en compte significative des phénotypes comportementaux dans une caractérisation plus large des capacités d'adaptation des animaux d'élevage.

Il y a également un élan pour le développement d'outils de suivi plus globaux et de méthodes d'analyses automatisées pour une mesure en continu de paramètres physiologiques, des comportements, des métabolites via la métabolomique et l'utilisation des biosenseurs, du suivi immunitaire, de la composition des tissus et des produits via le déploiement des techniques d'imagerie 3D ou de spectrométrie par infra-rouge. Si certaines technologies sont déjà bien développées (température, rythme cardiaque et respiratoire, imagerie 3D, spectres MIR et NIR), leur utilisation à grande échelle en condition d'élevage est encore limitée en raison de la taille des équipements et de la difficulté de collecte des données en continu. Pour les porcs et les volailles on s'oriente vers la mesure infra-rouge de groupes d'animaux et l'identification individuelle pour effectuer le suivi de la température corporelle.

Le phénotypage en continu peut aussi être une solution pour l'amélioration de la santé en élevage. Par exemple, le projet IC-F2F PigletDetect (INRA et IFIP, cofinancé par l'équipementier français d'élevage ASSERVA ; <http://www.ifip.asso.fr/fr/content/pigletdetect-dans-le-cadre-de-linstitut-carnot-france-futur-elevage>) vise à vérifier si une détection précoce des maladies est possible par le suivi individuel et en temps réel des consommations d'eau et d'aliment par les porcs permettant une prise en charge individualisée et plus précoce des maladies et une limitation de l'usage de médicaments.

6. Edition des génomes ?

Grâce au séquençage du génome porcin (Groenen *et al.*, 2012) et les avancées dans les outils de bio-informatique, de transgénèse et de nucléases ciblées (Carlson *et al.* 2012, Tan *et al.* 2013), il est possible



d'imaginer la modification des génomes des animaux afin d'améliorer leurs performances immunitaires en utilisant les données génétiques produites dans cette thèse pour identifier les variants causaux.

L'avènement de la méthode de CRISPR-Cas9 (pour *Clustered Regularly Interspersed Short Palindromic Repeats* en anglais) laisse entrevoir des possibilités d'édition ciblée du génome des animaux d'élevage. CRISPR-Cas9 (Marraffini, 2015) est un système bactérien adaptatif qui a permis le développement d'un outil d'ingénierie ou de modification des génomes après les méthodes ZFN et TALEN. CRISPR-Cas9 améliore les limites des autres méthodes : temps, coût, efficacité, maîtrise du nombre de copies, localisation précise des modifications du génome. En élevage, la méthodologie CRISPR-Cas9 pourrait permettre des transferts de variants favorables dans une population (race, lignée) vers une autre population. Cela pourrait être une piste pour améliorer la santé des animaux.

Cependant, CRISPR-Cas9 est soumis à l'insuffisance de nos connaissances, notamment sur la durabilité des résistances mono-ou oligogéniques. Il sera important de savoir sur quels gènes et quelles mutations travailler. Ainsi, les approches de caractérisation du déterminisme des phénotypes comme présentées au cours de cette thèse, seront plus pertinentes que jamais. De plus, de nouvelles réglementations seront nécessaires avec des considérations éthiques et sociétales, bien au-delà de la recherche et de la production, afin que la société donne un statut aux animaux à génome édités par CRISPR-Cas9 (animaux transgéniques ou non) et décide ou non de son acceptabilité.



Bibliographie

- Ackermann, M., Sikora-Wohlfeld, W., and Beyer, A. (2013). Impact of natural genetic variation on gene expression dynamics. *PLoS Genet.* *9*, e1003514.
- Aguirre-Gamboa, R., Joosten, I., Urbano, P.C.M., van der Molen, R.G., van Rijssen, E., van Cranenbroek, B., Oosting, M., Smeekens, S., Jaeger, M., Zorro, M., *et al.* (2016). Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits. *Cell Rep.* *17*, 2474–2487.
- Albers, R., Antoine, J.-M., Bourdet-Sicard, R., Calder, P.C., Gleeson, M., Lesourd, B., Samartín, S., Sanderson, I.R., Van Loo, J., Vas Dias, F.W., *et al.* (2005). Markers to measure immunomodulation in human nutrition intervention studies. *Br. J. Nutr.* *94*, 452–481.
- Alibert L. , Badouard B. , Legendre V. , Maupertuis F. , Calvar C. , Quinsac A. , Roinsard A. , Cresson C. *Caractériser les conditions de la mise en œuvre et du développement d'une production porcine française biologique « PorcBio »* Innovations Agronomiques 49 (2016), 33-48
- Al-Murrani, W.K., Al-Rawi, A.J., Al-Hadithi, M.F., and Al-Tikriti, B. (2006). Association between heterophil/lymphocyte ratio, a marker of “resistance” to stress, and some production and fitness traits in chickens. *Br. Poult. Sci.* *47*, 443–448.
- Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., and Hansson, I. (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* *263*, 1771–1774.
- Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E.K., Rivas, M.A., Pirinen, M., Gutierrez-Arcelus, M., Smith, K.S., Kukurba, K.R., *et al.* (2015). The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.* *25*, 927–936.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *J. Stat. Softw.* *67*.
- Baud A, Mulligan MK, Casale FP, Ingels JF, Bohl CJ, et al. (2017) Genetic Variation in the Social Environment Contributes to Health and Disease. *PLOS Genetics* *13*(1): e1006498.
- Benton, M.C., Lea, R.A., Macartney-Coxson, D., Carless, M.A., Göring, H.H., Bellis, C., Hanna, M., Eccles, D., Chambers, G.K., Curran, J.E., *et al.* (2013). Mapping eQTLs in the Norfolk Island genetic isolate identifies candidate genes for CVD risk traits. *Am. J. Hum. Genet.* *93*, 1087–1099.
- Bidanel, Milan. *La recherche de QTL à l'aide de marqueurs: Résultats chez le porc*. INRA Prod. Anim., 2000, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 223-228
- Boichard, D. (2015). Sélection génomique chez les bovins laitiers : les raisons de son succès. *Sélectionneur Fr.* *66* 11-20 2015.
- Boichard, D., Ducrocq, V., Croiseau, P., and Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *C. R. Biol.* *339*, 274–277.
- Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* *296*, 752–755.



- Brem, R.B., Storey, J.D., Whittle, J., and Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* *436*, 701–703.
- Brodin, P., Jojic, V., Gao, T., Bhattacharya, S., Angel, C.J.L., Furman, D., Shen-Orr, S., Dekker, C.L., Swan, G.E., Butte, A.J., *et al.* (2015). Variation in the human immune system is largely driven by non-heritable influences. *Cell* *160*, 37–47.
- Brown, C.D., Mangravite, L.M., and Engelhardt, B.E. (2013). Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* *9*, e1003649.
- Bryois, J., Buil, A., Evans, D.M., Kemp, J.P., Montgomery, S.B., Conrad, D.F., Ho, K.M., Ring, S., Hurles, M., Deloukas, P., *et al.* (2014). Cis and Trans Effects of Human Genomic Variants on Gene Expression. *PLOS Genet.* *10*, e1004461.
- Calder, P.C. (2005). Polyunsaturated fatty acids and inflammation. *Biochem. Soc. Trans.* *33*, 423–427.
- Calder, P.C. (2007). Immunological parameters: what do they mean? *J. Nutr.* *137*, 773S–80S.
- Calder, P.C., and Kew, S. (2002). The immune system: a target for functional foods? *Br. J. Nutr.* *88 Suppl 2*, S165–177.
- Carlson D. F., Tan W., Lillico S. G., Stverakova D., Proudfoot C., Christian M., *et al.* . (2012a). Efficient talen-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U.S.A.* *109*, 17382–17387.
10.1073/pnas.1211446109
- Casneuf, T., Van de Peer, Y., and Huber, W. (2007). In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation. *BMC Bioinformatics* *8*, 461.
- Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* *16*, 195.
- Chamberlain, A.J., Vander Jagt, C.J., Hayes, B.J., Khansefid, M., Maret, L.C., Millen, C.A., Nguyen, T.T.T., and Goddard, M.E. (2015). Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* *16*, 993.
- Chaussabel, D. (2015). Assessment of immune status using blood transcriptomics and potential implications for global health. *Semin. Immunol.* *27*, 58–66.
- Chaussabel, D., Pascual, V., and Banchereau, J. (2010). Assessing the human immune system through blood transcriptomics. *BMC Biol.* *8*, 84.
- Chen, C., Wei, R., Qiao, R., Ren, J., Yang, H., Liu, C., and Huang, L. (2012). A genome-wide investigation of expression characteristics of natural antisense transcripts in liver and muscle samples of pigs. *PLoS One* *7*, e52433.
- Chen, X., Hasan, M., Libri, V., Urrutia, A., Beitz, B., Rouilly, V., Duffy, D., Patin, É., Chalmond, B., Rogge, L., *et al.* (2015). Automated flow cytometric analysis across large numbers of samples and cell types. *Clin. Immunol. Orlando Fla* *157*, 249–260.
- Cho, J. (2015). The heritable immune system. *Nat. Biotechnol.* *33*, 608–609.



- Choi, I., Bao, H., Kommadath, A., Hosseini, A., Sun, X., Meng, Y., Stothard, P., Plastow, G.S., Tuggle, C.K., Reecy, J.M., *et al.* (2014). Increasing gene discovery and coverage using RNA-seq of globin RNA reduced porcine blood samples. *BMC Genomics* 15, 954.
- Clapperton, M., Bishop, S.C., Cameron, N.D., and Glass, E.J. (2005). Associations of acute phase protein levels with growth performance and with selection for growth performance in Large White pigs. *Anim. Sci.* 81.
- Clapperton, M., Glass, E.J., and Bishop, S.C. (2008). Pig peripheral blood mononuclear leucocyte subsets are heritable and genetically correlated with performance. *Animal* 2, 1575–1584.
- Clapperton, M., Diack, A.B., Matika, O., Glass, E.J., Gladney, C.D., Mellencamp, M.A., Hoste, A., and Bishop, S.C. (2009). Traits associated with innate and adaptive immunity in pigs: heritability and associations with performance under different health status conditions. *Genet. Sel. Evol.* 41, 54.
- Conde, L., Bracci, P.M., Richardson, R., Montgomery, S.B., and Skibola, C.F. (2013). Integrating GWAS and expression data for functional characterization of disease-associated SNPs: an application to follicular lymphoma. *Am. J. Hum. Genet.* 92, 126–130.
- Costa, V., Angelini, C., De Feis, I., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *J. Biomed. Biotechnol.* 2010, 853916.
- Crick, F. (1970). Molecular biology in the year 2000. *Nature* 228, 613–615.
- Crick, F.H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.
- Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., *et al.* (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* 47, 353–360.
- Davis, A.K., Maney, D.L., and Maerz, J.C. (2008). The use of leukocyte profiles to measure stress in vertebrates: a review for ecologists. *Funct. Ecol.* 22, 760–772.
- Dawson, H.D., Loveland, J.E., Pascal, G., Gilbert, J.G.R., Uenishi, H., Mann, K.M., Sang, Y., Zhang, J., Carvalho-Silva, D., Hunt, T., *et al.* (2013). Structural and functional annotation of the porcine immunome. *BMC Genomics* 14, 332.
- De Jager, P.L., Hacohen, N., Mathis, D., Regev, A., Stranger, B.E., and Benoist, C. (2015). ImmVar project: Insights and design considerations for future studies of “healthy” immune variation. *Semin. Immunol.* 27, 51–57.
- Debaere, O. (2016). Ecoantibio : first plan for the reduction of the risks of antibiotic resistance in veterinary medicine (2012-2016).
- Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y., and Pritchard, J.K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinforma. Oxf. Engl.* 25, 3207–3212.
- Delanoue, E., Dockès, A.C., and Chouteau, A. (2017). Une étude auprès de 2000 Français révèle de multiples attentes sur l'élevage.



- Denyer, M.S., Wileman, T.E., Stirling, C.M.A., Zuber, B., and Takamatsu, H.-H. (2006). Perforin expression can define CD8 positive lymphocyte subsets in pigs allowing phenotypic and functional analysis of natural killer, cytotoxic T, natural killer T and MHC un-restricted cytotoxic T-cells. *Vet. Immunol. Immunopathol.* *110*, 279–292.
- DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.* *8*, e1002600.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* *29*, 15–21.
- Duffy, D., Rouilly, V., Libri, V., Hasan, M., Beitz, B., David, M., Urrutia, A., Bisiaux, A., Labrie, S.T., Dubois, A., *et al.* (2014). Functional analysis via standardized whole-blood stimulation systems defines the boundaries of a healthy immune response to complex stimuli. *Immunity* *40*, 436–450.
- Edfors-Lilja, I., Wattrang, E., Marklund, L., Moller, M., Andersson-Eklund, L., Andersson, L., and Fossum, C. (1998). Mapping quantitative trait loci for immune capacity in the pig. *J. Immunol. Baltim. Md 1950* *161*, 829–835.
- Edsgård, D., Iglesias, M.J., Reilly, S.-J., Hamsten, A., Tornvall, P., Odeberg, J., and Emanuelsson, O. (2016). GeneiASE: Detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information. *Sci. Rep.* *6*, 21134.
- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., *et al.* (2008). Genetics of gene expression and its effect on disease. *Nature* *452*, 423–428.
- Ernst, C.W., and Steibel, J.P. (2013). Molecular advances in QTL discovery and application in pig breeding. *Trends Genet. TIG* *29*, 215–224.
- Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., *et al.* (2014). Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* *343*, 1246949.
- Fan, H., and Hegde, P.S. (2005). The transcriptome in blood: challenges and solutions for robust expression profiling. *Curr. Mol. Med.* *5*, 3–10.
- Fang X., Mou Y., Huang Z., Li Y., Han L., Zhang Y., Feng Y., Chen Y., Jiang X., Zhao W., Sun X., Xiong Z., Yang L., Liu H., Fan D., Mao L., Ren L., Liu C., Wang J., Li K., Wang G., Yang S., Lai L., Zhang G., Li Y., Wang J., Bolund L., Yang H., Wang J., Feng S., Li S., Du Y. (2012). The sequence and analysis of a Chinese pig genome. *Gigascience* *1*, 16.
- Fehrmann, R.S.N., Jansen, R.C., Veldink, J.H., Westra, H.-J., Arends, D., Bonder, M.J., Fu, J., Deelen, P., Groen, H.J.M., Smolonska, A., *et al.* (2011). Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* *7*, e1002197.
- Flori, L., Gao, Y., Laloë, D., Lemonnier, G., Leplat, J.-J., Teillaud, A., Cossalter, A.-M., Laffitte, J., Pinton, P., de Vaureix, C., *et al.* (2011). Immunity traits in pigs: substantial genetic variation and limited covariation. *PLoS One* *6*, e22717.



- Freeman, T.C., Ivens, A., Baillie, J.K., Beraldi, D., Barnett, M.W., Dorward, D., Downing, A., Fairbairn, L., Kapetanovic, R., Raza, S., *et al.* (2012). A gene expression atlas of the domestic pig. *BMC Biol.* *10*, 90.
- Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V.K., Weiler, J.E., O'Brien, P.J., and MacLennan, D.H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* *253*, 448–451.
- Gao, Y., Flori, L., Lecardonnel, J., Esquerré, D., Hu, Z.-L., Teillaud, A., Lemonnier, G., Lefèvre, F., Oswald, I.P., and Rogel-Gaillard, C. (2010). Transcriptome analysis of porcine PBMCs after in vitro stimulation by LPS or PMA/ionomycin using an expression array targeting the pig immune response. *BMC Genomics* *11*, 292.
- Gilad, Y., Rifkin, S.A., and Pritchard, J.K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet. TIG* *24*, 408–415.
- Goddard, M.E., and Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* *10*, 381–391.
- Göring, H.H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charlesworth, J., Cole, S.A., Jowett, J.B.M., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G., *et al.* (2007). Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* *39*, 1208–1216.
- Gourbeyre, P., Berri, M., Lippi, Y., Meurens, F., Vincent-Naulleau, S., Laffitte, J., Rogel-Gaillard, C., Pinton, P., and Oswald, I.P. (2015). Pattern recognition receptors in the gut: analysis of their expression along the intestinal tract and the crypt/villus axis. *Physiol. Rep.* *3*.
- Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., *et al.* (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* *491*, 393–398.
- Groeneveld, E., Kovac, M., and Mielenz, N. (2010). VCE User's Guide and Reference Manual Version 6.0.
- Grubbs, J.K., Dekkers, J.C.M., Huff-Lonergan, E., Tuggle, C.K., and Lonergan, S.M. (2016). Identification of potential serum biomarkers to predict feed efficiency in young pigs. *J. Anim. Sci.* *94*, 1482–1492.
- GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* *45*, 580–585.
- GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* *348*, 648–660.
- GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, *et al.* (2017). Genetic effects on gene expression across human tissues. *Nature* *550*, 204–213.
- Harvey, C.T., Moyerbrailean, G.A., Davis, G.O., Wen, X., Luca, F., and Pique-Regi, R. (2015). QuASAR: quantitative allele-specific analysis of reads. *Bioinforma. Oxf. Engl.* *31*, 1235–1242.
- Hasan, M., Beitz, B., Rouilly, V., Libri, V., Urrutia, A., Duffy, D., Cassard, L., Di Santo, J.P., Mottez, E., Quintana-Murci, L., *et al.* (2015). Semi-automated and standardized cytometric procedures for multi-panel and multi-parametric whole blood immunophenotyping. *Clin. Immunol. Orlando Fla* *157*, 261–276.



- Hasin-Brumshtein, Y., Hormozdiari, F., Martin, L., van Nas, A., Eskin, E., Lusi, A.J., and Drake, T.A. (2014). Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics* 15, 471.
- Haverson, K., Bailey, M., Stokes, C.R., Simon, A., LeFlufy, L., Banfield, G., Chen, Z., Hollemweguer, E., and Ledbetter, J.A. (2001). Monoclonal antibodies raised to human cells — specificity for pig leukocytes. *Vet. Immunol. Immunopathol.* 80, 175–186.
- Hayes, B. (2013). Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). *Methods Mol. Biol. Clifton NJ* 1019, 149–169.
- Hayes, B.J., Lewin, H.A., and Goddard, M.E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet. TIG* 29, 206–214.
- He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. *Science* 322, 1855–1857.
- IFIP, Institut du Porc (2016a). Le porc en chiffres.
- IFIP, Institut du Porc (2016b). Bilan d'activité.
- Jansen, R.C., and Nap, J.P. (2001). Genetical genomics: the added value from segregation. *Trends Genet. TIG* 17, 388–391.
- Joehanes, R., Zhang, X., Huan, T., Yao, C., Ying, S.-X., Nguyen, Q.T., Demirkale, C.Y., Feolo, M.L., Sharopova, N.R., Sturcke, A., *et al.* (2017). Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol.* 18, 16.
- Kadowaki, H., Suzuki, E., Kojima-Shibata, C., Suzuki, K., Okamura, T., Onodera, W., Shibata, T., and Kano, H. (2012). Selection for resistance to swine mycoplasmal pneumonia over 5 generations in Landrace pigs. *Livest. Sci.* 147, 20–26.
- Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., *et al.* (2005). Antisense transcription in the mammalian transcriptome. *Science* 309, 1564–1566.
- Kelly, S.A., Nehrenberg, D.L., Hua, K., Garland, T., and Pomp, D. (2014). Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. *Physiol. Genomics* 46, 593–601.
- Kendzioriski, C., and Wang, P. (2006). A review of statistical methods for expression quantitative trait loci mapping. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 17, 509–517.
- Kukurba, K.R., and Montgomery, S.B. (2015). RNA Sequencing and Analysis. *Cold Spring Harb. Protoc.* 2015, 951–969.
- Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nat. Rev. Genet.* 7, 385–394.
- Lagarrigue, S., Martin, L., Hormozdiari, F., Roux, P.-F., Pan, C., van Nas, A., ... Lusi, A. J. (2013). Analysis of Allele-Specific Expression in Mouse Liver by RNA-Seq: A Comparison With *Cis*-eQTL Identified Using Genetic Linkage. *Genetics*, 195(3), 1157–1166.



- Landmark-Høyvik, H., Dumeaux, V., Nebdal, D., Lund, E., Tost, J., Kamatani, Y., Renault, V., Børresen-Dale, A.-L., Kristensen, V., and Edvardsen, H. (2013). Genome-wide association study in breast cancer survivors reveals SNPs associated with gene expression of genes belonging to MHC class I and II. *Genomics* *102*, 278–287.
- Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A.C., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., *et al.* (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* *501*, 506–511.
- Larson, N.B., McDonnell, S., French, A.J., Fogarty, Z., Cheville, J., Middha, S., Riska, S., Baheti, S., Nair, A.A., Wang, L., *et al.* (2015). Comprehensively evaluating cis-regulatory variation in the human prostate transcriptome by using gene-level allele-specific expression. *Am. J. Hum. Genet.* *96*, 869–882.
- Lê Cao, K.-A., Martin, P.G., Robert-Granié, C., and Besse, P. (2009). Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* *10*, 34.
- Liew, C.-C., Ma, J., Tang, H.-C., Zheng, R., and Dempsey, A.A. (2006). The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J. Lab. Clin. Med.* *147*, 126–132.
- Liu, H., Nguyen, Y.T., Nettleton, D., Dekkers, J.C.M., and Tuggle, C.K. (2016). Post-weaning blood transcriptomic differences between Yorkshire pigs divergently selected for residual feed intake. *BMC Genomics* *17*.
- Liu, H., Smith, T.P.L., Nonneman, D.J., Dekkers, J.C.M., and Tuggle, C.K. (2017). A high-quality annotated transcriptome of swine peripheral blood. *BMC Genomics* *18*, 479.
- Lopes, M.S., Silva, F.F., Harlizius, B., Duijvesteijn, N., Lopes, P.S., Guimarães, S.E., and Knol, E.F. (2013). Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. *BMC Genet.* *14*, 92.
- Lunney, J.K. (2007). Advances in swine biomedical model genomics. *Int. J. Biol. Sci.* *3*, 179–184.
- MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., *et al.* (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* *335*, 823–828.
- Mach, N., Gao, Y., Lemonnier, G., Lecardonnell, J., Oswald, I.P., Estellé, J., and Rogel-Gaillard, C. (2013). The peripheral blood transcriptome reflects variations in immunity traits in swine: towards the identification of biomarkers. *BMC Genomics* *14*, 894.
- Mach, N., Berri, M., Estellé, J., Levenez, F., Lemonnier, G., Denis, C., Leplat, J.-J., Chevaleyre, C., Billon, Y., Doré, J., *et al.* (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* *7*, 554–569.
- Mackay, T.F.C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* *15*, 22–33.
- Mackay, T.F.C., Stone, E.A., and Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* *10*, 565–577.



- Mair, K.H., Sedlak, C., Käser, T., Pasternak, A., Levast, B., Gerner, W., Saalmüller, A., Summerfield, A., Gerds, V., Wilson, H.L., *et al.* (2014). The porcine innate immune system: An update. *Dev. Comp. Immunol.* *45*, 321–343.
- Maisonasse, P., Bouguyon, E., Piton, G., Ezquerra, A., Urien, C., Deloizy, C., Bourge, M., Leplat, J.-J., Simon, G., Chevalier, C., *et al.* (2016). The respiratory DC/macrophage network at steady-state and upon influenza infection in the swine biomedical model. *Mucosal Immunol.* *9*, 835–849.
- Mallard, B.A., and Wilkie, B.N. (2007). Phenotypic, genetic and epigenetic variation of immune response and disease resistance traits of pigs. *Adv. Pork Prod. Proc. Banff Pork Semin.*
- Mallard, B.A., Wilkie, B.N., Kennedy, B.W., and Quinton, M. (1992). Use of estimated breeding values in a selection index to breed Yorkshire pigs for high and low immune and innate resistance factors. *Anim. Biotechnol.* *3*, 257–280.
- Mallard, B.A., Emam, M., Paibomesai, M., Thompson-Crispi, K., and Wagter-Lesperance, L. (2015). Genetic selection of cattle for improved immunity and health. *Jpn. J. Vet. Res.* *63 Suppl 1*, S37-44.
- Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* *526*, 55–61.
- Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* *2*, 986–991.
- Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. *Hum. Mol. Genet.* *15*, R17–R29.
- Mehta, D., Heim, K., Herder, C., Carstensen, M., Eckstein, G., Schurmann, C., Homuth, G., Nauck, M., Völker, U., Roden, M., *et al.* (2013). Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood. *Eur. J. Hum. Genet. EJHG* *21*, 48–54.
- Menke, A., Rex-Haffner, M., Klengel, T., Binder, E.B., and Mehta, D. (2012). Peripheral blood gene expression: it all boils down to the RNA collection tubes. *BMC Res. Notes* *5*, 1.
- Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* *10*, 155–159.
- Meurens, F., Summerfield, A., Nauwynck, H., Saif, L., and Gerds, V. (2012). The pig: a model for human infectious diseases. *Trends Microbiol.* *20*, 50–57.
- Meuwissen, T.H., and Goddard, M.E. (2001). Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol. GSE* *33*, 605–634.
- Mills, J.D., Kawahara, Y., and Janitz, M. (2013). Strand-Specific RNA-Seq Provides Greater Resolution of Transcriptome Profiling. *Curr. Genomics* *14*, 173–181.
- Min, J.L., Taylor, J.M., Richards, J.B., Watts, T., Pettersson, F.H., Broxholme, J., Ahmadi, K.R., Surdulescu, G.L., Lowy, E., Gieger, C., *et al.* (2011). The use of genome-wide eQTL associations in lymphoblastoid cell lines to identify novel genetic pathways involved in complex traits. *PLoS One* *6*, e22070.
- Mohr, S., and Liew, C.-C. (2007). The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol. Med.* *13*, 422–432.



- Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* *430*, 743–747.
- Moyerbrailean, G.A., Richards, A.L., Kurtz, D., Kalita, C.A., Davis, G.O., Harvey, C.T., Alazizi, A., Watzka, D., Sorokin, Y., Hauff, N., *et al.* (2016). High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* *26*, 1627–1638.
- Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.-C., Freiman, A., Sams, A.J., Hebert, S., *et al.* (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* *167*, 657–669.e21.
- Neumaier, A., and Groeneveld, E. (1998). Restricted maximum likelihood estimation of covariances in sparse linear models. *Genet. Sel. Evol.* *30*, 3–26.
- Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermitzakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* *6*, e1000895.
- Orrù, V., Steri, M., Sole, G., Sidore, C., Viridis, F., Dei, M., Lai, S., Zoledziewska, M., Busonero, F., Mulas, A., *et al.* (2013). Genetic variants regulating immune cell levels in health and disease. *Cell* *155*, 242–256.
- Pai, A.A., Pritchard, J.K., and Gilad, Y. (2015). The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet.* *11*, e1004857.
- Palmer, C., Diehn, M., Alizadeh, A.A., and Brown, P.O. (2006). Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* *7*, 115.
- Pastinen, T. (2010). Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet.* *11*, 533–538.
- Pawelec G., Solana R., Remarque E., Mariani (1998) Impact of aging on innate immunity ; *J Leukoc Biol.* 1998 Dec;64(6):703-12.
- Peñagaricano, F., Valente, B.D., Steibel, J.P., Bates, R.O., Ernst, C.W., Khatib, H., and Rosa, G.J. (2015). Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. *BMC Syst. Biol.* *9*.
- Petretto, E., Mangion, J., Dickens, N.J., Cook, S.A., Kumaran, M.K., Lu, H., Fischer, J., Maatz, H., Kren, V., Pravenec, M., *et al.* (2006). Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet.* *2*, e172.
- Phocas, F., Belloc, C., Bidanel, J., Delaby, L., Dourmad, J.Y., Dumont, B., Ezanno, P., Fortun-Lamothe, L., Foucras, G., Frappat, B., *et al.* (2016). Review: Towards the agroecological management of ruminants, pigs and poultry through the development of sustainable breeding programmes: I-selection goals and criteria. *Anim. Int. J. Anim. Biosci.* *10*, 1749–1759.
- Pierce, B.L., Tong, L., Chen, L.S., Rahaman, R., Argos, M., Jasmine, F., Roy, S., Paul-Brutus, R., Westra, H.-J., Franke, L., *et al.* (2014). Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. *PLoS Genet.* *10*, e1004818.



- Ponsuksili, S., Zebunke, M., Murani, E., Trakooljul, N., Krieter, J., Puppe, B., Schwerin, M., and Wimmers, K. (2015). Integrated Genome-wide association and hypothalamus eQTL studies indicate a link between the circadian rhythm-related gene PER1 and coping behavior. *Sci. Rep.* *5*, 16264.
- Quach, H., Rotival, M., Pothlichet, J., Loh, Y.-H.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., *et al.* (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* *167*, 643–656.e17.
- Ramayo-Caldas, Y., Mach, N., Lepage, P., Levenez, F., Denis, C., Lemonnier, G., Leplat, J.-J., Billon, Y., Berri, M., Doré, J., *et al.* (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J.* *10*, 2973–2977.
- Ramos, A.M., Crooijmans, R.P.M.A., Affara, N.A., Amaral, A.J., Archibald, A.L., Beever, J.E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., *et al.* (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* *4*, e6524.
- Reiner, G., Eckert, J., Peischl, T., Bochert, S., Jäkel, T., Mackenstedt, U., Joachim, A., Dauschies, A., and Geldermann, H. (2002). Variation in clinical and parasitological traits in Pietrain and Meishan pigs infected with *Sarcocystis miescheriana*. *Vet. Parasitol.* *106*, 99–113.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* *43*, e47.
- Rockman, M.V., and Kruglyak, L. (2006). Genetics of global gene expression. *Nat. Rev. Genet.* *7*, 862–872.
- Rothkötter, H.-J. (2009). Anatomical particularities of the porcine immune system--a physician's view. *Dev. Comp. Immunol.* *33*, 267–272.
- Rotival, M., Zeller, T., Wild, P.S., Maouche, S., Szymczak, S., Schillert, A., Castagné, R., Deiseroth, A., Proust, C., Brocheton, J., *et al.* (2011). Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet.* *7*, e1002367.
- Salmon Henri ; Bull. Acad. Vet France – 2011 Tome 164 n°3 <http://www.academie-veterianire-defrance.org/> Transmission of maternal immunity in pigs and ruminants after birth, 26 mai 2011
- Sanchez, M.-P., Tribout, T., Iannuccelli, N., Bouffaud, M., Servin, B., Tenghe, A., Dehais, P., Muller, N., Del Schneider, M.P., Mercat, M.-J., *et al.* (2014). A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. *Genet. Sel. Evol. GSE* *46*, 12.
- Sasayama, D., Hori, H., Nakamura, S., Miyata, R., Teraishi, T., Hattori, K., Ota, M., Yamamoto, N., Higuchi, T., Amano, N., *et al.* (2013). Identification of single nucleotide polymorphisms regulating peripheral blood mRNA expression with genome-wide significance: an eQTL study in the Japanese population. *PLoS One* *8*, e54967.
- Satya, R.V., Zavaljevski, N., and Reifman, J. (2012). A new strategy to reduce allelic bias in RNA-Seq readmapping. *Nucleic Acids Res.* *40*, e127.
- Schadt, E.E., Monks, S.A., Drake, T.A., Lusk, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G., *et al.* (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* *422*, 297–302.



Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.

Schomberg D. T., Armando Tellez, Jennifer J. Meudt, Dane A. Brady, Krista N. Dillon, Folagbayi K. Arowolo, Joan Wicks, Serge D. Rousselle, Dhanansayan Shanmuganayagam (2016) Miniature Swine for Preclinical Modeling of Complexities of Human Disease for Translational Scientific Discovery and Accelerated Development of Therapies and Medical Devices *Toxicologic Pathology* Vol 44, Issue 3, pp. 299 - 314

Schramm, K., Marzi, C., Schurmann, C., Carstensen, M., Reinmaa, E., Biffar, R., Eckstein, G., Gieger, C., Grabe, H.-J., Homuth, G., *et al.* (2014). Mapping the genetic architecture of gene regulation in whole blood. *PLoS One* 9, e93844.

Schroyen, M., and Tuggle, C.K. (2015). Current transcriptomics in pig immunity research. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* 26, 1–20.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., *et al.* (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA N. Y. N* 22, 839–851.

Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., Chudin, E., Barker, D.L., Dickinson, T., *et al.* (2008). Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet.* 4, e1000006.

Seyednasrollah, F., Laiho, A., and Elo, L.L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* 16, 59–70.

Shen, H., and Huang, J.Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* 99, 1015–1034.

Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311.

Shimazu, T., Borjigin, L., Katayama, Y., Li, M., Satoh, T., Watanabe, K., Kitazawa, H., Roh, S., Aso, H., Kazuo, K., *et al.* (2014). Genetic selection for resistance to mycoplasmal pneumonia of swine (MPS) in the Landrace line influences the expression of soluble factors in blood after MPS vaccine sensitization. *Anim. Sci. J. Nihon Chikusan Gakkaiho* 85, 365–373.

Shin, H., Shannon, C.P., Fishbane, N., Ruan, J., Zhou, M., Balshaw, R., Wilson-McManus, J.E., Ng, R.T., McManus, B.M., Tebbutt, S.J., *et al.* (2014). Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS One* 9, e91041.

Siemelink, M.A., and Zeller, T. (2014). Biomarkers of coronary artery disease: the promise of the transcriptome. *Curr. Cardiol. Rep.* 16, 513.

Simon, L.M., Chen, E.S., Edelstein, L.C., Kong, X., Bhatlekar, S., Rigoutsos, I., Bray, P.F., and Shaw, C.A. (2016). Integrative Multi-omic Analysis of Human Platelet eQTLs Reveals Alternative Start Site in Mitofusin 2. *Am. J. Hum. Genet.* 98, 883–897.



- Siwek, M., Slawinska, A., Rydzanicz, M., Wesoly, J., Fraszczak, M., Suchocki, T., Skiba, J., Skiba, K., and Szyda, J. (2015). Identification of candidate genes and mutations in QTL regions for immune responses in chicken. *Anim. Genet.* *46*, 247–254.
- Skelly, D.A., Ronald, J., and Akey, J.M. (2009). Inherited variation in gene expression. *Annu. Rev. Genomics Hum. Genet.* *10*, 313–332.
- Smith, R.M., Webb, A., Papp, A.C., Newman, L.C., Handelman, S.K., Suhy, A., Mascarenhas, R., Oberdick, J., and Sadee, W. (2013). Whole transcriptome RNA-Seq allelic expression in human brain. *BMC Genomics* *14*, 571.
- Stefani, G., and Slack, F.J. (2008). Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell Biol.* *9*, 219–230.
- Strimbu, K., and Tavel, J.A. (2010). What are Biomarkers? *Curr. Opin. HIV AIDS* *5*, 463–466.
- Tan W, Carlson DF, Lancto CA, Garbe JR, Webster DA, Hackett PB, Fahrenkrug SC (2013) Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proc Natl Acad Sci U S A.* 2013 Oct 8; *110*(41):16526-31.
- Thomas, S., Rouilly, V., Patin, E., Alanio, C., Dubois, A., Delval, C., Marquier, L.-G., Fauchoux, N., Sayegrih, S., Vray, M., *et al.* (2015). The Milieu Intérieur study - an integrative approach for study of human immunological variance. *Clin. Immunol. Orlando Fla* *157*, 277–293.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinforma. Oxf. Engl.* *25*, 1105–1111.
- Tribout, T., Larzul, C., and Phocas, F. (2012). Efficiency of genomic selection in a purebred pig male line. *J. Anim. Sci.* *90*, 4164–4176.
- Tribout, T., Larzul, C., and Phocas, F. (2013). Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. *Genet. Sel. Evol. GSE* *45*, 40.
- Urrutia, A., Duffy, D., Rouilly, V., Posseme, C., Djebali, R., Illanes, G., Libri, V., Albaud, B., Gentien, D., Piasecka, B., *et al.* (2016). Standardized Whole-Blood Transcriptional Profiling Enables the Deconvolution of Complex Induced Immune Responses. *Cell Rep.* *16*, 2777–2791.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., *et al.* (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* *43*, 11.10.1-33.
- Visscher, A.H., Janss, L.L.G., Niewold, T.A., and de Greef, K.H. (2002). Disease incidence and immunological traits for the selection of healthy pigs. A review. *Vet. Q.* *24*, 29–34.
- Wang, X., Tang, H., Teng, M., Li, Z., Li, J., Fan, J., Zhong, L., Sun, X., Xu, J., Chen, G., *et al.* (2014). Mapping of hepatic expression quantitative trait loci (eQTLs) in a Han Chinese population. *J. Med. Genet.* *51*, 319–326.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* *10*, 57–63.



- Watson-Haigh, N.S., Kadarmideen, H.N., and Reverter, A. (2010). PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinforma. Oxf. Engl.* *26*, 411–413.
- Whitney, A.R., Diehn, M., Popper, S.J., Alizadeh, A.A., Boldrick, J.C., Relman, D.A., and Brown, P.O. (2003). Individuality and variation in gene expression patterns in human blood. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 1896–1901.
- Wilkie, B., and Mallard, B. (1999). Selection for high immune response: an alternative approach to animal health maintenance? *Vet. Immunol. Immunopathol.* *72*, 231–235.
- Williams, R.B.H., Chan, E.K.F., Cowley, M.J., and Little, P.F.R. (2007). The influence of genetic variation on gene expression. *Genome Res.* *17*, 1707–1716.
- Wimmers, K., Murani, E., Schellander, K., and Ponsuksili, S. (2009). QTL for traits related to humoral immune response estimated from data of a porcine F2 resource population. *Int. J. Immunogenet.* *36*, 141–151.
- Wood, D.L.A., Nones, K., Steptoe, A., Christ, A., Harliwong, I., Newell, F., Bruxner, T.J.C., Miller, D., Cloonan, N., and Grimmond, S.M. (2015). Recommendations for Accurate Resolution of Gene and Isoform Allele-Specific Expression in RNA-Seq Data. *PLoS One* *10*, e0126911.
- Wu, H., Gaur, U., Mekchay, S., Peng, X., Li, L., Sun, H., Song, Z., Dong, B., Li, M., Wimmers, K., *et al.* (2015). Genome-wide identification of allele-specific expression in response to *Streptococcus suis* 2 infection in two differentially susceptible pig breeds. *J. Appl. Genet.* *56*, 481–491.
- Würschum, T. (2012). Mapping QTL for agronomic traits in breeding populations. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* *125*, 201–210.
- Xia, W., Zhu, X.W., Mo, X.B. *et al.* *Hum Genet* (2017) 136: 451. <https://doi.org/10.1007/s00439-017-1771-1>
- Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., Liang, S., Pedersen, A.Ø., Kjeldsen, N.J., Liu, C., *et al.* (2016). A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* 16161.
- Yang, H., Huang, X., Fang, S., He, M., Zhao, Y., Wu, Z., Yang, M., Zhang, Z., Chen, C., and Huang, L. (2017). Unraveling the Fecal Microbiota and Metagenomic Functional Capacity Associated with Feed Efficiency in Pigs. *Front. Microbiol.* *8*, 1555.
- Yang, Y., Tang, Z., Fan, X., Xu, K., Mu, Y., Zhou, R., and Li, K. (2016). Transcriptome analysis revealed chimeric RNAs, single nucleotide polymorphisms and allele-specific expression in porcine prenatal skeletal muscle. *Sci. Rep.* *6*, 29039.
- Zhao, X.-M., and Qin, G. (2013). Identifying Biomarkers with Differential Analysis. In *Bioinformatics for Diagnosis, Prognosis and Treatment of Complex Diseases*, B. Shen, ed. (Dordrecht: Springer Netherlands), pp. 17–31.
- Zou, F., Chai, H.S., Younkin, C.S., Allen, M., Crook, J., Pankratz, V.S., Carrasquillo, M.M., Rowley, C.N., Nair, A.A., Middha, S., *et al.* (2012). Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* *8*, e1002707.



Annexe 1 : Les projets ANR IMMOPIG et SUS_FLORA

A. Le projet IMMOPIG

Le projet IMMOPIG a été financé pour mener une étude sur le déterminisme génétique de la réponse immunitaire chez le porc avec trois volets: le phénotypage de paramètres de la réponse immunitaire d'animaux de performances zootechniques connues, des analyses génétiques pour calculer les héritabilités et corrélations entre mesures, et enfin l'analyse du transcriptome des cellules sanguines d'animaux contrastés pour des mesures de la réponse immunitaire. Un total de 445 et 491 animaux a été prélevé dans les stations expérimentales du Rheu et du Magneraud, respectivement.

Grâce à ce projet, des héritabilités modérées à fortes pour de nombreux paramètres immuns ont été calculées et des corrélations génétiques significatives entre paramètres immuns et zootechniques identifiées, dont un grand nombre sont négatives. Une puce à ADN enrichie en gènes de l'immunité a été conçue. L'analyse du transcriptome sanguin des animaux montre que ce phénotype est informatif pour une partie des caractères immuns mesurés. Ce projet a ainsi permis de confirmer l'existence d'un contrôle génétique de nombreux paramètres immuns et l'intérêt d'inclure des phénotypes moléculaires pour affiner leur caractérisation.

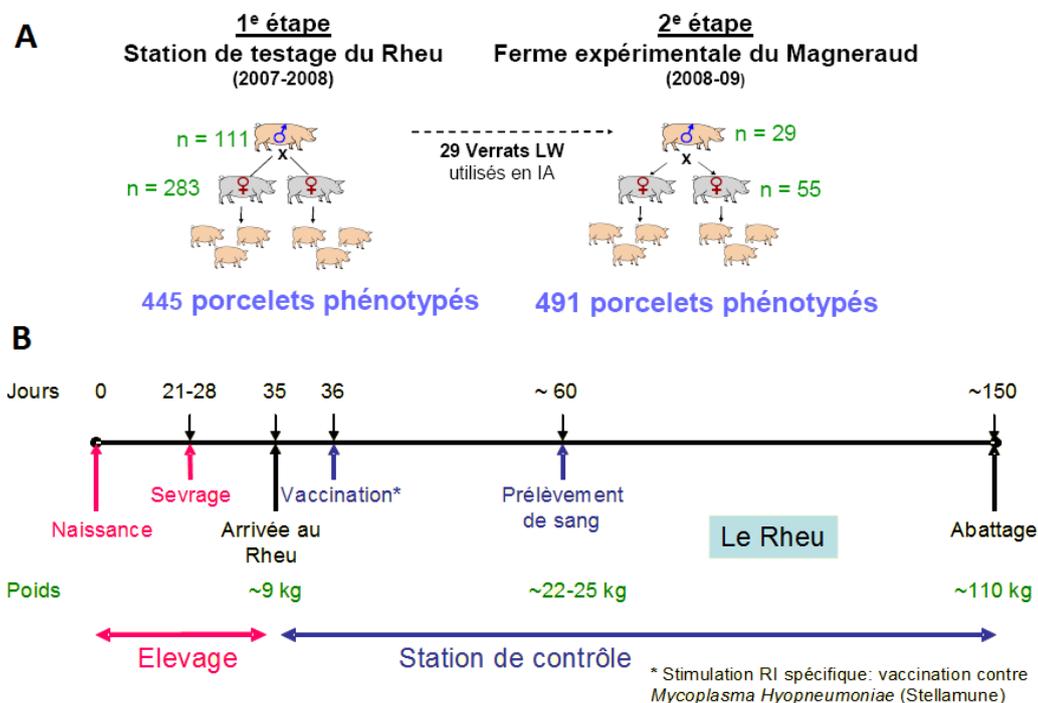


Figure A40 : A) Protocole de production des animaux du projet global ; B) Etapes de production des animaux élevés au Rheu et utilisés dans cette thèse.



Deux groupes d'animaux ont été produits dans le cadre de ce projet, impliquant deux stations expérimentales (Figure A1A) : la station expérimentale du Rheu et la station expérimentale du Magneraud. Les animaux pour lesquels des données ont été utilisés dans ce projet de thèse (Etude 1 pour l'analyse ASE) étaient des animaux issus de la station expérimentale du Rheu. Ces animaux étaient produits en élevage, sevrés à 21-28 jours et amenés dans la station expérimentale à 35 jours. Puis ils ont été vaccinés pour *Mycoplasma hyopneumoniae* à 36 jours. Les prélèvements de sang ont été effectués à 60 jours.

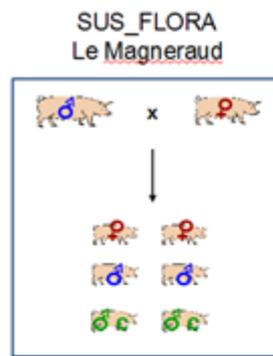
B. Le projet SUS_FLORA

Le projet SUS_FLORA avait pour but d'inclure le microbiote intestinal comme un phénotype à part entière à intégrer dans les travaux sur la capacité immune chez le porc et l'analyse de l'architecture génétique des paramètres de réponse immunitaire. Le projet visait d'une part l'étude simultanée de la variabilité du génome de l'hôte et du métagénome du microbiote intestinal, en lien avec sa capacité immunitaire, et d'autre part l'analyse des interactions locales entre le microbiote et l'épithélium intestinal. Connaître les facteurs qui influent sur la composition du microbiote et bien évaluer son impact dans un modèle porc contribueront à identifier des leviers d'action en élevage et à améliorer les connaissances chez les mammifères monogastriques (incluant l'homme). Le projet était basé sur l'étude de 560 porcelets de race Large White vaccinés contre *Mycoplasma Hyopneumoniae*, issus de 90 familles. Du sang et des fèces ont été prélevés sur les porcelets âgés de 60 jours. Les trois études présentées dans ce travail de thèse ont utilisé des données issues de ce projet.

Ce projet a mis en évidence que le microbiote intestinal se stabilise après 36 jours et répartit les animaux en deux groupes. A 60 jours, ces deux groupes sont confirmés et leur microbiote diffère par des abondances significativement différentes de certains genres bactériens. Une héritabilité significative pour l'abondance des genres bactériens les plus fréquents a été mise en évidence. Des covariations ont été identifiées entre composition du microbiote et des paramètres immunitaires et de croissance. Il a également été établit une carte d'expression de l'ensemble des transcrits et de gènes d'intérêt (TLR, NOD, MHC, cytokines) dans des tissus cibles.



A



B

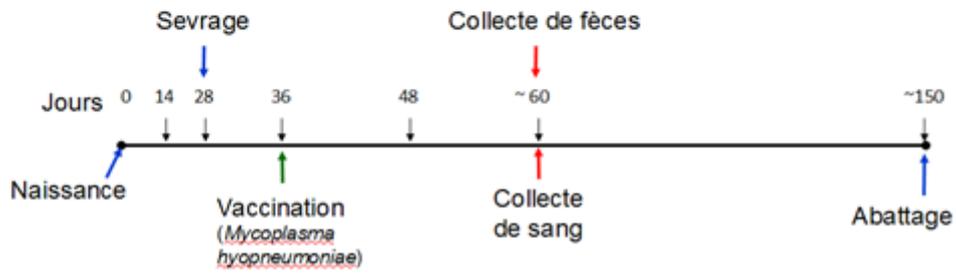


Figure A41 : A) Animaux produits dans le projet SUS_FLORA; B) Protocole de production des animaux du projet SUS_FLORA

Annexe 2 : Caractères immunitaires porcins mesurables dans le sang

		Code	Dénomination complète	Unité	Méthode de mesures	Fonctions/interprétations	Valeurs
Phénotypes descriptifs	Dosages sériques	CRP	Protéine C réactive	mg/mL	Dosage ELISA	Protéine synthétisée par les hépatocytes, marqueur précoce de la réponse inflammatoire.	
		HAPT O	Haptoglobine	mg/mL	Dosage	Protéine de synthèse hépatique (demi-vie 2-5 jours). Diminution en cas d'hémolyse (intra- et extra- vasculaire) et d'insuffisance hépatocellulaire. Augmente en cas d'inflammation subaiguë ou chronique.	
		IgA	Immunoglobuline A	mg/mL	Dosage	Immunoglobuline présente dans les sécrétions. Rôle essentiel de défense contre les agressions externes.	
		IgG_M H	Immunoglobuline G anti- <i>Mycoplasma hyopneumoniae</i>	E/P	Dosage	Reflet de l'efficacité de la vaccination réalisée en moyenne 30 à 35 jours plus tôt.	
	Sous-types de leucocytes	NK	Cellules NK	% des PBMC	Cytométrie en flux, CD3-CD8α+ CD16+	Détruisent les cellules infectées par une action cytotoxique médiée par la perforine et sécrètent des cytokines	
		NKT	lymphocytes T tueurs		Cytométrie en flux CD3+CD8α+ CD16+	Lymphocytes à la limite des immunités innée et adaptative. Reconnait des lipides et glycolipides présentés par la molécule CD1d.	
		gdT	lymphocytes T γδ		Cytométrie en flux CD3+γδTCR+	Participent à l'immunité innée et adaptative. Chez l'homme ils représentent 1 à 5% des cellules du sang, chez certaines espèces animales dont le porc, ils sont beaucoup plus nombreux.	
		CD4+CD8-	Lymphocytes T auxiliaires ou LT helper		Cytométrie en flux, CD3+γδTCR- CD4+CD8α-	Aident à la régulation des fonctions lymphocytaires par sécrétion de cytokines.	
		CD4+CD8+	Lymphocytes T auxiliaires mémoires		Cytométrie en flux, CD3+γδTCR- CD4+CD8α+	Spécifiques de l'antigène qu'elles ont rencontré en premier et peuvent être mises en œuvre lors de la réponse immunitaire secondaire.	



		CD4- CD8+	Lymphocytes T cytotoxiques		Cytométrie en flux CD3+ γδTCR- CD4- CD8α+	Les lymphocytes T cytotoxiques sont responsables de l'immunité cellulaire aboutissant à la mort de la cellule cible. On observe une libération des granules cytotoxiques (lysosomes particuliers) qui contiennent deux catégories de molécules que l'on appelle des cytotoxines : la perforine et les sérine-estérases.	
		CD4- CD8-	Lymphocytes T null		Cytométrie en flux CD3+ γδTCR- CD4- CD8α-	Lymphocytes T immatures	
		cDC1	Cellules dendritiques conventionnelles de type 1		Cytométrie en flux, CD172a+ CD14- CMH II high CD163-	Les CD sont des cellules présentatrices d'antigènes dites « professionnelles ». Elles sont capables de capturer et d'apprêter l'antigène pour le présenter aux LT. Les cDC1 induisent une réponse polarisée Th1 alors que les cDC2 une réponse Th2.	
		cDC2	Cellules dendritiques conventionnelles de type 2		Cytométrie en flux, CD172a+ CD14- CMH II high CD163+		
		CD14+ CD163 -	Monocyte CD14+ CD163-		Cytométrie en flux, CD172a+ CD14+ CD163-	Cellules du sang qui évoluent en macrophages ou cellules dendritiques en passant dans les tissus. Chez le porc, 4 sous-types de monocytes sont identifiés dans le sang selon les marqueurs CD14 et CD163 et correspondent à des stades de maturation différents : (I) CD14+ CD163+, (II) CD14- CD163-, (III) CD14+ CD163+, (IV) CD14- CD163-.	
		CD14- CD163 -	Monocyte CD14- CD163-		Cytométrie en flux, CD172a+ CD14- MHC II low CD163-		
		CD14+ CD163 +	Monocyte CD14+ CD163+		Cytométrie en flux, CD172a+ CD14+ CD163+		
		CD14- CD163 +	Monocyte CD14- CD163+		Cytométrie en flux, CD172a+ CD14- MHC II low CD163+		
		GB	Globules Blancs	10 ³ /mm ³	Hémo- cytomètre	Augmentation en cas d'infection	11-22
		LYM	Lymphocytes	%		Diminution en cas d'infection systémique sévère, augmentation en cas d'infections bactériennes sévères (après 2 ou 3 jours), d'inflammation, de stress, d'exercice physique intense	35-64



	Numé- ration de formule sanguine	MON	Monocytes			Augmentation en cas d'infection chronique, de maladie virale, d'infection parasitaire, de récupération d'une infection aiguë	2-10
		GRA	Granulocytes			Diminution en cas d'infection virale ou bactérienne toxémique. Augmentation en cas d'infection virale aiguë ou bactérienne, de désordre inflammatoire chronique et de stress	28-62
		EOS	Eosinophiles			Augmentation en cas d'infections parasitaires, de désordres inflammatoires	0-11
		GR	Globules Rouges	10 ⁶ /mm ³		Diminution en cas d'hémorragie, de déficiences alimentaires. Augmentation en cas de déshydratation, de diarrhée	5-8
		Hgb	Hémoglobine	g/dL	10-17		
		Hct	Hématocrite	%	32- 50		
		VGM	Volume Globulaire Moyen	fL			50-68
		TCMH	Teneur Corpusculaire Moyenne en Hémoglobine	pg		Diminution en cas d'anémie liée à une carence en fer ou des désordres inflammatoires chroniques. Augmentation en cas de déficiences nutritionnelles (vitamine B12 ou folate), maladie du foie, hypothyroïdisme.	17 -21
		CCMH	Concentration Corpusculaire Moyenne en Hémoglobine	g/dL			27-40
		PLT	Plaquettes	10 ³ /mm ³		Diminution en cas d'infection virale. Augmentation en cas de carence en fer et d'anémie hémolitique	250-750
GRA/ LYM	Ratio Granulocytes/Lymphocytes	Unité universelle	calcul	Ratio classiquement utilisé en écologie pour caractériser le statut immunitaire des animaux. Diminue en cas de stress, d'effort, d'infection.			
Phénotypes de réponse	Dosage de cytokines et chimiokines sécrétées suite à une	TNF α	Facteur de nécrose de tumeur alpha (TNF-α)	pg/mL	Dosage multiplex CBA (Cytokine Bead Assay) OU ELISA	Produit par les cellules épithéliales, les cellules infectées, les neutrophiles, les macrophages, les cellules NK. Impliqué dans l'inflammation systématique, chronique et la réaction aiguë, il active la voie de NF-κB, l'apoptose et la nécrose programmée.	
		IL2	Cytokine IL2			Produite par les lymphocytes T helper pour stimuler la prolifération et l'activation de cellules T, B et NK	
		IL1b	Cytokine IL1β			Produite par les cellules épithéliales, les cellules infectées, les neutrophiles, les macrophages activées. Médiateur important de la réaction inflammatoire. Active la prolifération cellulaire, la différenciation et l'apoptose et induit la fièvre.	



stimulation (durée variable) in vitro sur sang total	IL8	Chimiokine IL8			Produite par les cellules épithéliales, endothéliales et innées. Assure le recrutement des neutrophiles sur le site de l'infection et leur activation.	
	IL17	Cytokine IL17			Cytokine pro-inflammatoire. Recrute et active les neutrophiles. Principalement produite les lymphocytes Th17.	
	INFg	Interféron gamma			Produit par les cellules NK, les lymphocytes T gd, les lymphocytes Th1, les lymphocytes T cytotoxiques. Rôle principal dans l'immunité à médiation cellulaire contre les pathogènes intracellulaires par activation de macrophages, cellules NK, CTL, maintien de la différenciation Th1. Indique une infection virale	
	IL10	Interleukine 10			Produit par les macrophages polarisés M2, les lymphocytes T reg , les lymphocytes Th2. Activité anti-inflammatoire et immunosuppressive. Inhibe IL2, TNF et certains interférons et module différentes cellules.	
Capacités de phagocytose	Phagocytose	% de cellules phagocytaires	% des leucocytes		Proportion de cellules capables de phagocyter la bactérie <i>E coli</i> opsonisée	
	Phago_MFI	Moyenne d'intensité de fluorescence	MFI		Quantité de bactéries <i>E coli</i> internalisées par les cellules phagocytaires	
	Granulo_Phago	% de granulocytes phagocytaires	% des granulocytes		Proportion de granulocytes capables de phagocyter la bactérie <i>E coli</i> opsonisée	
	Granulo_Phago_MFI	Moyenne d'intensité de fluorescence	MFI		Quantité de bactéries <i>E coli</i> internalisées par les granulocytes phagocytaires	
	Mono_Phago	% de monocytes phagocytaires	% des monocytes		Proportion de monocytes capables de phagocyter la bactérie <i>E coli</i> opsonisée	
	Mono_Phago_MFI	Moyenne d'intensité de fluorescence	MFI		Quantité de bactéries <i>E coli</i> internalisées par les monocytes phagocytaires	
				Cytométrie en flux		



Publications et Communications

Posters

Doctorate School Day 2015 (Gif-sur-Yvette, France) : An eGWAS analysis of the porcine whole blood transcriptome ; T. Maroilley, G. Lemonnier, J. Lecardonnel, M. Moroldo, C. Rogel-Gaillard and J. Estellé

ISAFG 2015 (Piacenza, Italy) : An eGWAS ananalysis of the porcine whole blood transcriptome ;T.Maroilley, M. Ballester, G. Lemonnier, M-J. Mercat, Y. Billon, M. Moroldo, C. Rogel-Gaillard and J. Estellé

Journées de la Recherche Porcine 2016 (Paris, France) : Etude de la capacité immunitaire chez le porc Large White : recherche de marqueurs génétiques liés à l'expression des gènes dans le sang ; T. Maroilley, M. Ballester, G. Lemonnier, M-J. Mercat, Y. Billon, M. Moroldo, C. Rogel-Gaillard and J. Estellé

ISAG 2017 (Dublin, Ireland) : RNA-Seq analysis of gut-associated lymphoid tissue in pigs revealed few differences in transcription profiles of ileal and jejunal Peyer's patches ; T. Maroilley, D.Esquerré, C. Chevaleyre, G. Lemonnier, J-J. Leplat, S. Vincent-Naulleau, M-J. Mercat, Y. Billon, P. Lepage, C. Rogel-Gaillard and J. Estellé

Communications orales

ISAG 2016 (Salt Lake City, USA): Allele specific expression analysis of the porcine blood transcriptome reveals extensive cis-regulation in immunity-related genes ; T. Maroilley, G. Lemonnier, D. Esquerré, M-J. Mercat, C. Rogel-Gaillard and J. Estellé (communication + ePoster) ; Prix Poster

Doctorate School Day 2016 (Versailles, France) and Animal Genetic Department Seminar 2016 (Toulouse, France): Genetic control of gene blood expression in pigs ; T. Maroilley, J. Estellé and C. Rogel-Gaillard

Publications

Etude de la capacité immunitaire chez le porc Large White: recherche de marqueurs génétiques liés à l'expression des gènes dans le sang : T. Maroilley, M. Ballester, G. Lemonnier, M.J. Mercat, Y. Billon, M Moroldo, C. Rogel-Gaillard and J. Estellé ; JRP 2016

Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis ; T. Maroilley,



G.Lemonnier, J. Lecardonnell, D. Esquerré, Y. Ramayo-Caldas, M.J. Mercat, C. Rogel-Gaillard and J.Estellé ; BMC Genomics (accepté le 25/11/2017)

Immunome differences between porcine ileal and jejunal Peyer's patches revealed by global transcriptome sequencing of gut-associated lymphoid tissues ; T. Maroilley, M.Berri, C.Chevaleyre, S. Mélo, F. Meurens, D. Esquerré, G. Lemonnier, J-L. Coville, J-J. Leplat, A. Rau, B. Bed'hom, Y.Billon, S. Vincent-Naulleau, M-J. Mercat, C. Rogel-Gaillard and J. Estellé ; Scientific Reports (soumis)

Production

Guide pour la soumission de projet ITN des Actions Marie-Curie, H2020 lors d'une mission doctorale au sein de la cellule Europe de l'Université Paris-Saclay



Title : Genetics and Genomics of the immune capacities in pigs : eQTL approach and study of blood biomarkers

Keywords : pigs, genetics, genomics, immunity, biomarkers

Abstract : A better understanding of the mechanisms for resistance to pathogens along with the characterization of the immune capacity has become a priority for research in breeding. The overall objective of this PhD project was to use phenotyping, genetic and genomic information to study the genetic architecture of the immune capacity in the pig and to contribute to the identification of genetic markers and blood biomarkers that predict variations of immune parameter levels. The study was focused on three complementary axes and the results obtained were based on the use of data collected as part of the IMMOPIG and SUS_FLORA projects financed by the ANR, for which 450 and 560 piglet cohorts were sampled at 60 days of age, three weeks after vaccination against *Mycoplasma hyopneumoniae*.

We analyzed the genetic determinism of gene expression in the blood using an eGWAS (60K SNP genotyping and blood transcriptome for 242 animals) confirmed in part by an allele specific expression (ASE) of transcripts (RNA-Seq on 38 animals). The eGWAS results showed multiple local (n=2839) and distant associations between the SNP polymorphisms and transcription variations, spread over all the chromosomes. The ASE analyses confirmed the *cis* genetic control of gene expression, with allele regulation being found for 763 genes. The biological functions associated were notably associated with RNA processing and immunity. The region of the major histocompatibility complex was found particularly rich in eQTL signals and genes with ASE in the blood.

We studied the co-variations between immune parameters and blood transcriptomes for 243 individuals.

We studied the co-variations between immune parameters and blood transcriptome for 243 individuals. The immune parameters included the blood count, cell subpopulation characterization by flow cytometry, serum assays (reactive C protein, haptoglobin, antibodies specific for *Mycoplasma hyopneumoniae*), immune response after *in vitro* stimulation of peripheral blood (phagocytosis, IL1b, IL2, IL6, IL8, IL10, TNFa, INFg cytokines). We confirmed the heritability of numerous immune parameters and identified covariations with gene profiles, providing hypotheses on biomarker candidates.

We also led a functional analysis on four animals at 70 days-of-age in order to characterize and compare the transcriptome profiles of peripheral blood and three gut-associated lymphoid tissues (mesenteric lymph nodes, jejunal and ileal Payer's patches). The RNA-Seq data showed differential expression between tissues; this number was more limited between the two types of Peyer's patches. Interestingly, among the biological functions enriched by the differentially expressed genes between the Peyer's patches, we identified the Th1 and Th2 lymphocyte differentiation pathways, which was in agreement with an overabundance of B lymphocytes in the ileal Peyer's patches.

Together these results provide new information on the understanding of the genetic determinism of immune parameter variations in the pig, the search for causal polymorphisms of these variations and the identification of relevant blood biomarkers for phenotyping immune competence.



Titre : Génétique et Génomique de la capacité immunitaire chez le porc : approche eQTL et étude de biomarqueurs sanguins

Mots clés : porc, génétique, génomique, immunité, biomarqueurs

Résumé : Une meilleure compréhension des mécanismes de résistance aux agents pathogènes couplée à une caractérisation de la capacité immune devient un axe de recherche prioritaire en élevage. L'objectif global de la thèse est d'exploiter des informations de phénotypage, de génétique et de génomique pour étudier l'architecture génétique de la capacité immune chez le porc et contribuer à l'identification de marqueurs génétiques et de biomarqueurs sanguins prédictifs de variations de niveaux de paramètres immuns. Le projet s'articule autour de trois axes complémentaires et les résultats obtenus sont basés sur l'exploitation des jeux de données issus des projets IMMOPIG et SUS_FLORA financés par l'ANR, pour lesquels des cohortes de 450 et 560 porcelets ont été prélevées à 60 jours d'âge, trois semaines après une vaccination contre *Mycoplasma hyopneumoniae*.

Nous avons analysé le déterminisme génétique de l'expression des gènes dans le sang par une analyse eGWAS (génotypage 60K SNP et transcriptome du sang pour 242 animaux) validée pour partie par une étude de la régulation spécifique d'allèles (ASE) des transcrits (RNA-Seq sur 38 animaux). Les résultats d'eGWAS mettent en évidence de multiples associations locales (n=2839) et à distance (n=1752) entre des polymorphismes de SNP et des variations de transcription, réparties sur l'ensemble des chromosomes. Les analyses ASE confirment l'importance du contrôle génétique en *cis*, avec une régulation allélique trouvée pour 763 gènes. Les fonctions biologiques associées sont notamment reliées au processing des ARN et à l'immunité. La région du complexe majeur d'histocompatibilité a été trouvée particulièrement riche en eQTL et ASE.

L'ensemble de ces données a permis d'établir, pour le porc, une première cartographie des eQTL et gènes soumis à ASE dans le sang.

Nous avons étudié les co-variations entre des paramètres immunitaires et le transcriptome du sang pour 243 individus.

Les paramètres immunitaires incluent la numération formule sanguine, la caractérisation de sous-populations cellulaires par cytomètre de flux, des dosages sériques (protéine C réactive, haptoglobine, anticorps spécifiques anti *Mycoplasma hyopneumoniae*), des dosages de réponse suite à des stimulations *in vitro* du sang total (phagocytose, cytokines IL1b, IL2, IL6, IL8, IL10, TNFa, INFg). Nous avons confirmé l'héritabilité de nombreux paramètres immuns et identifié des covariations avec des profils géniques, offrant des hypothèses sur des biomarqueurs candidats.

Nous avons également conduit une analyse fonctionnelle sur quatre animaux de 70 jours d'âge pour caractériser et comparer les profils de transcrits du sang et de trois tissus lymphoïdes associés au tube digestif (ganglion mésentérique, plaques de Peyer jéjunales et iléales). Les données RNA-Seq ont mis en évidence des différentiels d'expression entre les tissus, ce nombre étant plus limité entre les deux types de plaques de Peyer. De manière tout à fait intéressante, parmi les fonctions biologiques enrichies par les gènes différentiellement exprimés entre les deux plaques de Peyer, sont identifiées les voies de différenciation lymphocytaires Th1 et Th2, en cohérence avec une surabondance de lymphocytes B dans les plaques de Peyer iléales.

L'ensemble de ces résultats apporte des informations nouvelles pour avancer dans la compréhension du déterminisme génétique des variations de paramètres immunitaires chez le porc, la recherche de polymorphismes causaux de ces variations et l'identification de biomarqueurs sanguins pertinents pour phénotyper la compétence immunitaire.

