



HAL
open science

Génomique du muscle porcin et survie néonatale

Laurence Liaubet

► **To cite this version:**

Laurence Liaubet. Génomique du muscle porcin et survie néonatale. Life Sciences [q-bio]. INPT, 2019. <tel-04795730>

HAL Id: tel-04795730

<https://hal.inrae.fr/tel-04795730v1>

Submitted on 21 Nov 2024

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

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



HAL Authorization



**Habilitation à diriger des recherches
Délivrée par l'Institut National Polytechnique de Toulouse**

Présentée et soutenue par

Laurence LIAUBET

13 septembre 2019

Génomique du muscle porcin et survie néonatale

**Unité de Recherche : UMR1388 Génétique, Physiologie et Systèmes
d'Élevage (GenPhySE, INRA)**

Jury

Rapporteurs

André Almeida	Professeur, Institut Supérieur d'Agronomie de l'Université de Lisbonne
Sandrine Lagarrigue	Professeure, Agrocampus Ouest, ENSAR
Claire Rogel-Gaillard	Directrice de Recherche, INRA

Examineurs

Hervé Rémignon	Professeur, INPT-ENSAT
Nathalie Viguerie	Chargée de Recherche, Inserm/Université Paul Sabatier

Table des matières

Liste des figures	3
Liste des tables	4
Liste des abréviations	5
Partie 1 : Activités de recherche : Bilan et perspectives	7
A. Avant-propos	7
B. Contexte	7
C. Génomique du muscle porcin.....	9
1. Mise en place de la maturité musculaire à la naissance.....	9
a. Importance du muscle dans l'étude des performances de production et des caractères fonctionnels.	9
b. Génomique du muscle en fin de gestation	10
c. Analyse du transcriptome musculaire en fin de gestation chez trois espèces d'intérêt agronomique - Diversité intra espèce vs inter-espèces.....	17
d. Architecture nucléaire et régulation de l'expression des gènes musculaires en fin de gestation.....	20
2. Diversité génétique fonctionnelle	24
3. Etude d'un caractère complexe par une approche intégrative : la tendreté de la viande	28
a. Recherche de gènes régulés en fonction d'un phénotype quantitatif.	28
b. Intégration des données issues de l'analyse du transcriptome et du protéome. .	31
c. Organisation génomique des gènes régulés.....	32
d. Régulation génétique de l'expression (eQTL).	34
e. Recherche de gènes prédicteurs du phénotype.....	42
D. Perspectives à court et moyen terme : Survie néonatale, intégration de données multi-omiques, multi tissus, prédiction à la naissance et intégration génétique.	46
1. Intégration des données des fœtus en fin de gestation relatives à la survie néonatale.....	48
a. Trois jeux de données métabolomiques; plasma, urine et liquide amniotique...	49
b. Les six transcriptomes et deux protéomes.....	51
c. Données physiologiques.....	53
d. Données morphologiques fœtales et chez le nouveau-né.....	55
2. Etude des interactions fœto-maternelles	56
3. Prédiction de la survie et de la qualité de survie à la naissance.....	58
4. Intégration génétique - génomique	63
Références	66
Partie 2. Curriculum Vitae, Liste des publications, Activités d'enseignement et Responsabilités collectives.....	74

A. Curriculum Vitae	74
Fonction à l'INRA.....	74
Management	74
Formation	74
Formations complémentaires	74
Travaux scientifiques	75
Enseignement - autres expériences professionnelles	76
B. Liste des publications	76
C. Encadrements	79
D. Responsabilités collectives.....	82
1. Fonctionnement des collectifs.....	82
Au sein du laboratoire	82
Au sein de l'INRA	82
Sur le site toulousain pour diverses animations scientifiques	82
Réseaux européens	82
2. Animation de l'équipe GENOROBUST, Génétique des systèmes en lien avec l'adaptation et la Robustesse	83

Liste des figures

Figure 1 : Travaux antérieurs à mon arrivée à l'INRA.	6
Figure 2 : Schéma simplifiant les étapes de formation du muscle au cours de la gestation chez le porc.	10
Figure 3 : Remaniement de l'expression des 12326 transcrits (Bonferroni 1%) entre 90 et 110 jdg dans le muscle fœtal.	10
Figure 4 : Illustration des variations observées de la glycolyse et de la gluconéogenèse dans le muscle fœtal à 90 et 110 jdg.	11
Figure 5 : Coévolution de la myosine embryonnaire, de la myosine adulte (rapide) et du glycogène musculaire en fin de gestation.	14
Figure 6 : Evolution des chaînes lourdes de myosines chez le porc.	14
Figure 7 : Analyse PLS entre les données musculaires du transcriptome et du protéome à 110 jdg.	15
Figure 8 : Réseau à 110 jdg prenant en compte les relations intra-jeu de données transcriptome, protéome et les phénotypes (réseaux de co-expression) et entre les différents niveaux d'information (réseaux bipartites).	16
Figure 9 : Vue d'ensemble de la gestation chez trois espèces (bovin, mouton et porc).	17
Figure 10 : Profils d'expression des six sets de gènes représentant (A) le cycle cellulaire, (B) la matrice extracellulaire, (C) les protéines des fibres rapides, (D) le transport mitochondrial, (E) le métabolisme lipidique et (F) le métabolisme du glycogène.	19
Figure 11 : Présentation du processus itératif d'inférence des graphes de co-expression autour d'IGF2.	21
Figure 12 : Sous-réseau 3 du réseau de co-expression et co-localisation nucléaire des allèles exprimés.	22
Figure 13 : Analyse de la diversité porcine exprimée à différents niveaux d'expression du génome.	24
Figure 14 : Qualité de prédiction de 22 phénotypes de production à partir du métabolome plasmatique.	25
Figure 15 : Représentation synthétique de l'implication de la voie des PI3Kinases dans différents processus cellulaires.	26
Figure 16 : Représentation de réseaux inférés spécifiques de chaque race à partir des 123 gènes les plus différenciellement exprimés chez les Large White et les Landrace.	27
Figure 17 : Présentation du dispositif expérimental pour l'analyse du transcriptome du muscle dans un contexte génétique.	29
Figure 18 : Expression de deux gènes (SELW et la TTN) en fonction des valeurs de forces de cisaillement mesurées sur viande cuite.	29
Figure 19 : Régression linéaire entre les valeurs d'expression et la force de cisaillement de huit gènes obtenues par deux méthodes différentes (microarray vs. qPCR).	30
Figure 20 : Réseau bibliographique obtenu avec le logiciel Ingenuity incluant 23 gènes différentiels essentiellement impliqués dans le développement et la morphologie du tissu musculaire.	31
Figure 21 : Analyse des données du protéome et du transcriptome du muscle Longissimus par une approche de réseaux de co-occurrence (logiciel IPA).	32
Figure 22 : Co-localisation entre des gènes différenciellement exprimés et des QTL pour le tendreté de la viande.	33
Figure 23 : Schéma résumant deux stratégies d'intégration de données d'expression et génétiques dans le cadre d'un réseau européen d'analyse des données transcriptome EADGENE (Jouffe et al. 2009).	34
Figure 24 : 335 régions génomiques (eQTL) identifiées pour 272 transcrits.	35
Figure 25 : Cartographie eQTL sur le chromosome 1 (A) et corrélogramme entre les gènes localisant entre 133-149 cM sur le chromosome 1 (B).	37
Figure 26 : Représentation simplifiée du réseau de co-expression des 272 gènes régulés par au moins un eQTL.	38
Figure 27 : Présentation détaillée du cluster 4.	39
Figure 28 : Réseau bibliographique comprenant 10 gènes reconnus par IPA et reliés au pH ultime.	40
Figure 29 : Première recherche de possibles marqueurs de qualité par calcul d'un score discriminant sur l'ensemble des valeurs de force de cisaillement.	42
Figure 30 : Application de la méthode des Forêts aléatoires pour identifier les gènes les plus discriminants du phénotype force de cisaillement (N > 30).	43
Figure 31 : Variations d'expression pour le meilleur prédicteur de la tendreté chez le porc selon le statut au polymorphisme identifié dans ce transcrit (données non publiées).	44
Figure 32 : Le gène MAGEL2 identifié comme prédicteur du stade et du génotype dans le muscle fœtal identifié par la méthode des forêts aléatoires.	45

Figure 33 : Déterminisme précoce des phénotypes tardifs et intégration génétique-génomique.	46
Figure 34 : Schéma expérimental du projet PORCINET.	48
Figure 35 : Analyse discriminante par OPLS-DA des profils métaboliques plasmatiques par RMN ¹ H réalisée avec 397 variables identifiées comme informatives.	50
Figure 36 : Premières analyses statistiques des données des transcriptomes de cinq tissus et du sang total.	52
Figure 37 : Analyse MFA réalisée avec les 15 gènes les plus différentiels pour l'interaction génotype et âge gestationnel et pour les trois tissus : muscle, intestin et sang total.	53
Figure 38 : A gauche. Glycogène musculaire chez 459 fœtus.	54
Figure 39 : A droite. Déplétion des réserves énergétiques dans les heures qui suivent la naissance.	54
Figure 40 : Exemple du gène IGF2 d'expression préférentiellement paternelle dans le muscle fœtal.	57
Figure 41 : Photo de six porcelets LW nouveau-nés.	59
Figure 42 : Exemple de deux gènes surexprimé ou sous-exprimé que chez les mâles LW morts (LW_M_M) dans les 3 jours après la naissance.	59
Figure 43 : Présentation des résultats de la sPLS pour le groupe de phénotypes relatif à la mobilité entre la naissance et 7 jours après.	60
Figure 44 : Réseau IPA obtenu avec les données des métabolites plasmatiques en fin de gestation.	61
Figure 45 : Mise en perspective des projets PORCINET et SuBPig.	62
Figure 46 : Graphe obtenu par inférence simple après détection de QTL et de pQTL et test de la pléiotropie.	64

Liste des tables

Table 1 : Six sets de gènes comparés entre et intra espèces.	18
Table 2 : Cinq tissus fœtaux analysés par étude du transcriptome et deux tissus par protéomique.	51

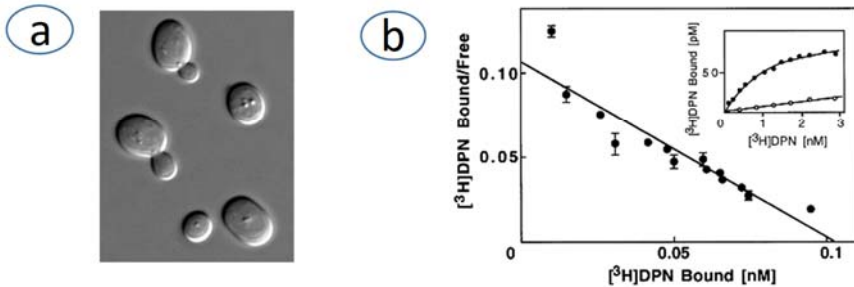
Liste des abréviations

ACP	Analyse en composantes principales
AK1	<i>Adénylate Kinase 1</i>
ANR	Agence nationale de la recherche
ARG2	<i>Arginase 2</i>
BTS	Brevet de Technicien Supérieur
CCA	analyse canonique des corrélations ou <i>canonical-correlation analysis</i>
CAH	Classification Ascendante Hiérarchique
CKMT2	<i>Creatine Kinase, Mitochondrial 2</i>
cM	centimorgan
CMJR	Consommation Moyenne Journalière Résiduelle
CTNNB1	<i>Catenin Beta 1</i>
DCN	<i>Decorin</i>
DEA	Diplôme d'Etudes Approfondies
DEG	<i>Differentially Expressed Genes</i>
DEP	<i>Differentially Expressed Probes</i>
DLK1	<i>Delta like non-canonical Notch ligand 1</i>
DUT	<i>Deoxyuridine Triphosphatase</i>
eQTL	<i>Expression quantitative trait loci</i>
EAPP	<i>2F Associated Phosphoprotein</i>
FDR	<i>False Discovery Rate</i>
GDF8 ou MSTN	<i>Growth Differentiation Factor 8</i> ou <i>Myostatin</i>
GenPhySE	Génétique Physiologie et Systèmes d'Élevage
GGM	<i>Gaussian Graphical Models</i>
GPI	<i>Glucose Phosphate Isomerase</i>
FDR	<i>False Discovery Rate</i> , procédure de contrôle du taux de fausses découvertes dans les situations où le nombre de tests statistiques est très important.
FISH	<i>3D fluorescent in situ hybridization</i>
HiC	<i>High-throughput Chromosome conformation capture</i> ou capture de conformation chromosomique sur l'ensemble du génome
HNRNPA1	<i>Heterogeneous Nuclear Ribonucleoprotein A</i>
IFIP	Institut du porc
IGF2	<i>Insulin like growth factor 2</i>
IPA	<i>Ingenuity Pathway Analysis</i> (QIAGEN Bioinformatics)
Jdg	jours de gestation
LDHA	<i>Lactate Dehydrogenase A</i>
LGC	Laboratoire de Génétique Cellulaire
LRT	<i>Likelihood Ratio Test</i>
MAGEL2	<i>MAGE Family Member L2</i>
MDS	<i>Multidimensional Scaling</i> ou Positionnement multidimensionnel
MEG3	<i>Maternally expressed 3</i>
MEST	<i>Mesoderm Specific Transcript</i>
MI-MFA	<i>Multiple Imputation-Multiple Factor Analysis</i>
MYH3	<i>Myosin heavy chain 3</i>
MyHC	<i>Myosin heavy chain</i> , myosine de chaîne lourde (famille de plusieurs protéines)
MYOD1	<i>Myogenic Differentiation 1</i>
O-PLS-DA	<i>Orthogonal Partial Least-Squares Discriminant Analysis</i>
OSP	Organisation de Sélection Porcine (partenaires privés qui s'occupent de sélection génétique)
PCIT	<i>Partial Correlation and Information Theory</i>
PCK2	<i>Phosphoenolpyruvate Carboxykinase 2, Mitochondrial</i>
PCR	<i>Polymerase Chain Reaction</i>
PGK1	<i>Phosphoglycerate Kinase 1</i>
PHKG1	<i>Phosphorylase Kinase Catalytic Subunit Gamma 1</i>
PLS et sPLS	<i>Partial Least Squares</i> ou régression par les moindres carrés partiels, sparse PLS
qPCR	PCR quantitative dite en temps réel ou <i>real-time qPCR</i>
QTL	<i>Quantitative Trait Loci</i> , locus de caractères quantitatifs
RF	<i>Random Forest</i> ou Forêts aléatoires
RMN	Résonance Magnétique Nucléaire
RPL32	<i>ribosomal protein L32</i>
RNF2	<i>Ring Finger Protein 2</i>
SSCP	<i>single strand conformation polymorphism</i> ou polymorphisme de conformation des acides nucléiques simple brin
SEPW1	<i>Selenoprotein W</i>
TASC	Tissu Adipeux Sous-Cutané
TGFβ	<i>Transforming Growth Factor</i>
TTN	<i>Titin</i>

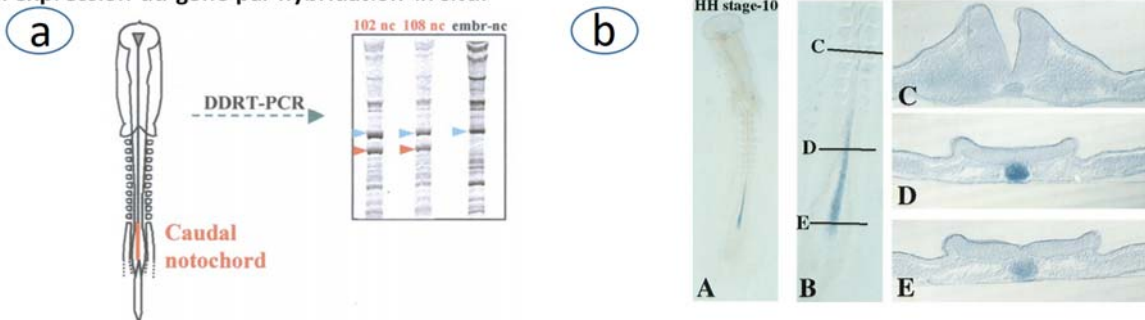
A. Interactions plantes-microorganismes : mise au point d'un modèle d'étude entre *A. thaliana* et *R. solanacearum* (Thèse de Pharmacie, LIPM/INRA). La photo a montre l'effet de l'infection à droite. La photo b montre, sur une coupe transversale, la colonisation des vaisseaux par la bactérie transformée (GUS) dans les vaisseaux de xylème (Deslandes et al., 1998).



B. Expression du récepteur humain mu aux opioïdes chez *S. cerevisiae* et Sf9/baculovirus (DEA Pharmacologie Cellulaire et Moléculaire, IPBS/CNRS ; Gaibelet et al., 1999).



C. a. Identification d'une métalloprotéase par differential display exprimée dans la notochorde caudale dans l'embryon de poulet (Thèse en Sciences, CBD/CNRS ; Liaubet et al., 2000) . b. Visualisation de l'expression du gène par hybridation in situ.



D. Caractérisation phénotypique de *MTM1*^{-/-}, gène impliqué dans une myopathie myotubulaire (stage post-doctoral, IGBMC/INSERM et I2MC/INSERM). Les photos a et b représentent des coupes transversales de muscles. A gauche les fibres musculaires normales, au milieu celles d'une souris *MTM1*^{-/-} (fibres plus petites). La photo c représente le phénotype *MTM1*^{-/-} chez un modèle unicellulaire.

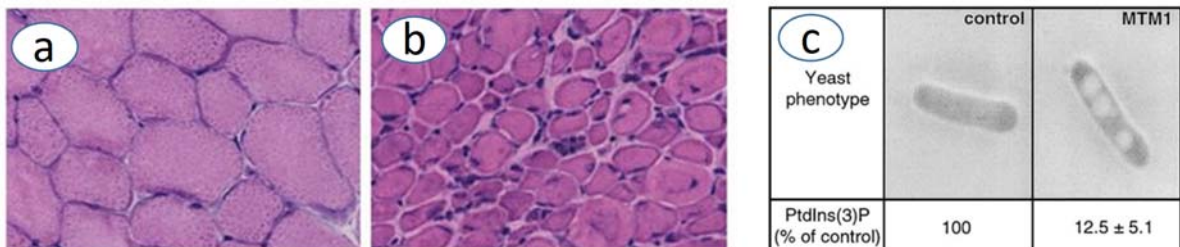


Figure 1 : Travaux antérieurs à mon arrivée à l'INRA. Ces travaux illustrent la variété des modèles et thématiques abordées autour de la question de l'identification et la caractérisation de gènes d'intérêt pour un phénotype.

Partie 1 : Activités de recherche : Bilan et perspectives

A. Avant-propos

Dans ce manuscrit, je vais reprendre 18 ans de recherches effectuées à l'INRA de Toulouse. Toutefois, mes premiers pas dans des laboratoires de recherches datent de 1991. Je venais de réussir le concours d'entrée en deuxième année de Pharmacie et je voulais absolument connaître le milieu de la Recherche. Je ne vais pas lister 10 ans de stages, mon doctorat de Pharmacie, le DEA et la thèse en Sciences. Mais je ne peux pas résister au plaisir d'illustrer la variété des thématiques abordées, des modèles vivants manipulés, des techniques utilisées et de quelques résultats dans la Figure 1.

Avant ma thèse de Docteur en Pharmacie, j'avais déjà effectué plusieurs stages en lien avec les modules que j'ai suivis en parallèle de mes études de pharmacie pour obtenir une Maîtrise des Sciences Biologiques et Médicales (Botanique et Biotechnologie Végétale, Ciblage Thérapeutique, Innovation thérapeutique) ou bien encore dans le cadre de mon externat hospitalier (Pharmacocinétique, Pharmacovigilance et Pharmacologie clinique en cancérologie).

La Figure 1 reprend les travaux de recherche ayant abouti à mes six premières publications (Deslandes *et al.* 1998; Godiard *et al.* 1998; Gaibelet *et al.* 1999; Liaubet *et al.* 2000; Laporte *et al.* 2002; Tronchère *et al.* 2004). Deux publications (Figure 1A) pour ma thèse de Pharmacie réalisée en génétique et biologie moléculaire végétale (Deslandes *et al.* 1998; Godiard *et al.* 1998) à l'INRA (Laboratoire d'Interactions Plantes Micro-organismes), une (Figure 1B) pour mon DEA de Pharmacologie Cellulaire et Moléculaire à l'Institut de Pharmacologie et Biologie Structurale (Gaibelet *et al.* 1999), une (Figure 1C) pour ma thèse en Sciences (Liaubet *et al.* 2000) et deux (Figure 1D) pour mon stage post-doctoral (Laporte *et al.* 2002; Tronchère *et al.* 2004).

B. Contexte

A mon arrivée à l'INRA dans le Laboratoire de Génétique Cellulaire en 2001, ma thématique concernait l'étude de critères de qualité des produits et la tendreté de la viande en particulier. Ce projet était dénommé D116 et avait pour objectif la recherche de marqueurs précoces de la qualité de la viande. Ce projet présentait des aspects assez novateurs, proposant d'analyser le transcriptome et le protéome du muscle porcin dans un dispositif génétique développé pour identifier des QTL. La partie transcriptome du muscle *Longissimus* m'a été confiée. Ce projet ne pouvait pas être réalisé sans le développement d'un outil dédié à l'analyse globale de l'expression des gènes (développement des *microarrays*) et des méthodologies associées (d'abord statistiques). L'investissement dans ce projet a été peu productif en termes de publications (Lobjois *et al.* 2008; Liaubet *et al.* 2011; Villa-Vialaneix *et al.* 2013). Par contre, ces débuts ont été très formateurs sur la question d'intégrer de l'information de génomique fonctionnelle avec des données de génétique. Les premières publications alliant transcriptome et génétique (eQTL) datent de 2002 (Brem *et al.* 2002; Schadt *et al.* 2003), avec une revue annonciatrice en 2001 (Jansen & Nap 2001) et des travaux français fondateurs intégrant déjà des données protéomiques et génétiques en 1988 (de Vienne *et al.* 1988).

Mes premières recherches avaient donc pour objectif d'analyser un caractère complexe (la tendreté de la viande), caractère d'autant plus complexe que les animaux n'avaient pas été choisis pour ne représenter que les valeurs extrêmes pour le phénotype. Bien au contraire, les animaux à analyser avaient été produits pour représenter autant que possible la variabilité phénotypique individuelle afin de pouvoir associer de l'information génétique.

La complexité d'analyse de telles données a été l'occasion de développer une collaboration passionnante avec deux statisticiennes : Magali San Cristobal (de mon Unité, GenPhySE) et Nathalie Vialaneix (Unité de Mathématiques et Informatique Appliquées de Toulouse, MIAT) avec 19 publications avec au moins l'une des deux comme co-auteur ainsi qu'un éditorial (Villa-Vialaneix *et al.* 2011) et un chapitre

d'ouvrage (Villa-Vialaneix et al. 2016).

Au cours des 10 premières années, j'ai encadré un étudiant en thèse pendant une année (Pierre Ferré, soutenance en 2004), une stagiaire post-doctorale (Valérie Lobjois, 2005-2006), la partie d'analyse du transcriptome d'un autre stagiaire post-doctoral (Dominique Hazard, en 2006 dans le cadre d'une collaboration avec Pierre Mormède du laboratoire PsyNuGen à Bordeaux), un seul étudiant niveau M2 et 11 étudiants de différents niveaux (BTS, DUT, Licence, M1) en biologie et en statistique. Les étudiants en statistique ont toujours été co-encadrés par Magali San Cristobal ou Nathalie Vialaneix.

Les 10 premières années ont donc été consacrées à l'analyse du muscle prélevé sur l'animal en croissance ou après abattage. L'intégration de l'information génétique s'est donc faite en co-localisant des gènes régulés avec des QTL connus pour les phénotypes liés à la tendreté de la viande (Lobjois *et al.* 2008), par l'identification de eQTL (Liaubet *et al.* 2011) mais aussi en étudiant la diversité génétique de l'expression des gènes en comparant plusieurs lignées porcines (projet ANR DeLiSus coordonné par Denis Milan et Magali San Cristobal, (SanCristobal et al. 2015)).

Fin des années 2000, on notera une évolution dans les questions sociétales et les questions venant de la filière professionnelle porcine. La question du bien-être animal et de la robustesse des animaux se fait plus pressante. J'ai été alors plus particulièrement interpellée sur la question de la survie des porcelets à la naissance et j'en ai fait une nouvelle thématique de recherche avec toujours le muscle comme tissu d'intérêt central. J'ai alors coordonné un projet ANR de 2010 à 2014, PORCINET " Approche intégrée de la maturité des porcelets " avec l'analyse des porcelets en fin de gestation et à la naissance. Ce projet a généré une quantité de données très importante toujours en cours d'analyse. Le challenge à relever est, depuis, d'analyser, d'intégrer et d'interpréter des données multi-omiques et multi-tissus.

Donc pour les projets de recherche sur le muscle squelettique (*longissimus*), je soulignerai que celui-ci a donc été étudié :

- Chez le fœtus en fin de gestation (Voillet et al. 2014; Marti-Marimon et al. 2018; Voillet et al. 2018) avec la thèse de Valentin Voillet,
- En régénération, après avoir créé une lésion par injection intramusculaire (Ferre et al. 2007) dans le cadre de la thèse de Pierre Ferré,
- Après abattage, avec l'étude de la tendreté de la viande et autres critères de composition ou de qualité mesurés sur des porcs obtenus par croisement (Lobjois *et al.* 2008; Liaubet *et al.* 2011; Villa-Vialaneix *et al.* 2013) avec le soutien de Valérie Lobjois en stage post-doctoral,
- Après abattage, avec l'étude de la diversité génétique de l'expression par comparaison de plusieurs lignées pures/parentales (SanCristobal *et al.* 2015).

Depuis 2010, mes approches intégratives se sont complexifiées. Il ne s'agit plus de s'intéresser au tissu musculaire seul mais à l'animal dans son ensemble. L'objectif est maintenant d'utiliser des approches multi-omiques et multi-tissus, de la fin de gestation à l'abattoir, en combinant des données phénotypiques toujours plus diversifiées comme des données d'analyses sanguines, des critères de vitalité, de réponses de stress voire, dernièrement, des données longitudinales de thermographie. Ces travaux sont possibles car nous profitons de la diversité génétique disponible dans nos Unités Expérimentales, Unité GenESI ; en particulier en utilisant des lignées génétiques très différentes (Large White, LW, et Meishan, MS, et les croisés réciproques) ainsi que des lignées divergentes pour l'efficacité alimentaire et les réponses de stress.

- L'intégration des données -omiques périnatales obtenues dans différents fonds génétiques fait l'objet de la thèse de Gaëlle Lefort débutée en septembre 2018.

Et pour finir, l'ensemble des études que j'ai réalisées depuis mon recrutement m'ont particulièrement sensibilisée à la question du bien-être et de la robustesse de nos animaux d'élevage, de leur capacité d'adaptation. Un des critères de robustesse qui nous intéresse tout particulièrement est l'adaptation au milieu extra-utérin que constitue la naissance. Ce qui fait échos à l'idée que l'élaboration précoce des phénotypes impacte la croissance, la santé et le bien-être des animaux. Dans ce cadre, mon projet de recherche interrogera la périnatalité pour identifier des leviers génétiques et génomiques en fin de

gestation avec un focus particulier sur la naissance avec l'évaluation de la possibilité de modéliser les capacités de survie des nouveau-nés.

C. Génomique du muscle porcin

Dans ce chapitre, je ne décrirai pas le déroulement de mes recherches de façon chronologique mais plutôt dans un sens développemental en discutant d'abord des mécanismes biologiques du muscle en fin de gestation puis du muscle prélevé à l'abattoir. L'évolution de certains phénotypes a été recherchée pour obtenir une carcasse moins grasse et plus musclée. En contrepartie, des phénotypes peuvent avoir été impactés négativement après des décennies de sélection génétique ; par exemple, la thermorégulation et la motricité à la naissance.

1. Mise en place de la maturité musculaire à la naissance.

Après une petite introduction (Chapitre 1.a), je présenterai les résultats basés sur les travaux de la thèse de Valentin Voillet et qui correspondent à deux de ses publications (Voillet *et al.* 2014; Voillet *et al.* 2018). La directrice de thèse de Valentin était Magali San Cristobal, j'étais co-directrice. Ce travail de thèse s'est inséré dans le réseau de collaboration mis en place dans le cadre du projet ANR PORCINET. En particulier, nous avons profité de l'expertise de Louis Lefaucheur, physiologiste spécialiste du muscle porcin dans l'Unité PEGASE du centre INRA Bretagne-Normandie, Rennes.

a. Importance du muscle dans l'étude des performances de production et des caractères fonctionnels.

La masse musculaire représente 35 à 60 % de la masse corporelle d'un porc destiné à la production de viande. Dans des conditions optimales de nutrition, la croissance de la masse musculaire est déterminée par le nombre, le type et la taille des fibres musculaires à la naissance. La formation des fibres musculaires est biphasique (Figure 2, (Foxcroft *et al.* 2006)). La formation des fibres secondaires a lieu entre 54 et 90 jours de gestation. Ensuite les fibres musculaires vont se développer par hypertrophie. Le nombre de fibres musculaires est déterminé par des facteurs génétiques et par l'environnement à des périodes critiques pour le développement musculaire (Rehfeldt *et al.* 2000). Les muscles sont ensuite utilisés pour la locomotion mais aussi dans la régulation métabolique. En effet, les muscles servent de réservoir de protéines et de glycogène. Les réserves protéiques peuvent être mobilisées pour subvenir au besoin de tout l'organisme en période de stress physiologique ou de maladie. Le muscle joue aussi un rôle important dans l'homéostasie du glucose avec la capacité du muscle (et du foie) à stocker du glycogène et à le mobiliser si besoin (van der Lende *et al.* 2001). La sélection génétique a porté sur une croissance maigre et cette croissance va dépendre du nombre de fibres et de leurs tailles ; donc va dépendre du nombre de fibres formées avant la naissance. De plus, il est décrit qu'une meilleure qualité de la viande est plutôt corrélée à un nombre de fibres élevé de taille modérée (Rehfeldt *et al.* 2000). La sélection génétique pour favoriser certaines performances peut avoir affecté certains organes et tissus entraînant une démesure dans la croissance au niveau cellulaire et tissulaire. Une croissance disproportionnée correspondrait à une répartition inadéquate des ressources (Huber 2018). Selon la théorie de l'allocation des ressources, les ressources (réserves énergétiques, nutriments) doivent être partagées entre la croissance, la maintenance de l'homéostasie, la santé, la production et la reproduction pendant toute la durée de la vie. Les animaux sélectionnés pour des performances de production (croissance maigre, prolificité) voient leur robustesse diminuer avec une plus grande sensibilité aux maladies, une locomotion affaiblie et des problèmes de comportement (Rauw *et al.* 1998; Rauw 2016).

b. Génomique du muscle en fin de gestation

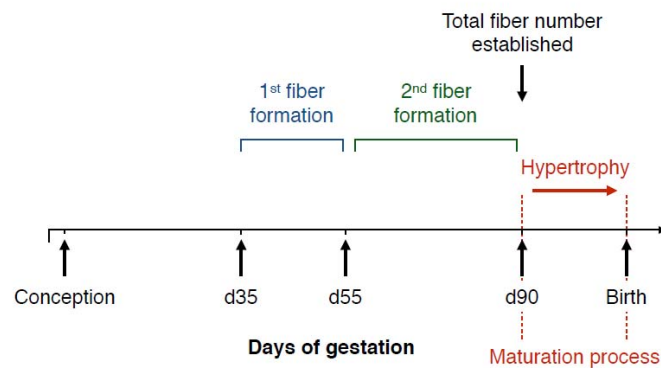


Figure 2 : Schéma simplifiant les étapes de formation du muscle au cours de la gestation chez le porc. Cette figure est inspirée de Foxcroft et al.(2006).

Le muscle squelettique joue un rôle important à la naissance en devant assurer une motricité immédiate et permettre au porcelet nouveau-né d'accéder aux tétines. Le muscle permet aussi d'assurer la thermorégulation du porcelet via la régulation de son métabolisme énergétique et le frissonnement (Herpin et al. 2002). L'objectif de l'étude était de caractériser les différences biologiques en comparant deux stades de fin du développement fœtal (Figure 2 et 3). Des truies de deux races, Large White (LW, race européenne dont les porcelets présentent une augmentation de la mortalité néonatale ces dernières décennies) et Meishan (MS, race chinoise dont les porcelets sont plus robustes), ont été croisées avec des pères des deux types génétiques (mélange de semence). Les fœtus ont été prélevés par césarienne au début et à la fin du dernier tiers de gestation (90 et 110 jours ; la naissance étant à 114 jours). Nous avons ainsi obtenu des fœtus « purs » et des croisés réciproques. Ces croisements permettent donc de comparer des génotypes extrêmes mais aussi d'évaluer le rôle respectif des deux demi-génomés LW ou MS selon qu'ils viennent du père ou de la mère. Le transcriptome a été obtenu sur 64 individus (8 par condition ; deux stades et 4 génotypes). L'analyse des données a été réalisée par Valentin Voillet dans le cadre de sa thèse.

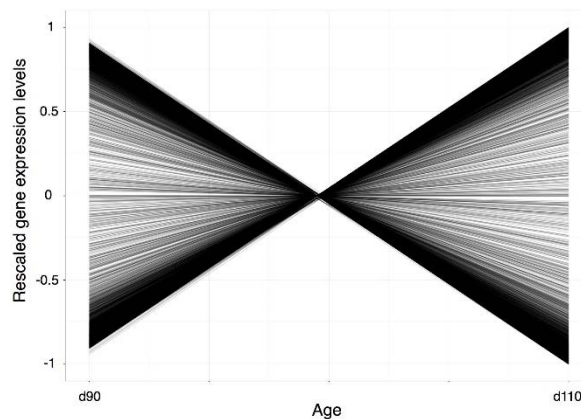


Figure 3 : Remaniement de l'expression des 12326 transcrits (Bonferroni 1%) entre 90 et 110 jdg dans le muscle fœtal.

L'analyse de l'expression différentielle entre les deux stades de fin de gestation (Figure 3) montre une modification majeure de l'expression génique avec plus de 12326 sondes différentielles (Bonferroni 1%) entre 90 et 110 jdg et pour les 4 génotypes. Avec la nouvelle annotation du génome porcin (P. Dehais, Sigenae, décembre 2018), pour ces 12326 sondes, 80% sont annotées et correspondent à 5167 gènes uniques. Cette variation drastique de l'expression génique avait déjà été observée chez le mouton (Sudre et al. 2003; Byrne et al. 2010) avec 2471 gènes différentiellement exprimés entre la fin de gestation et la naissance.

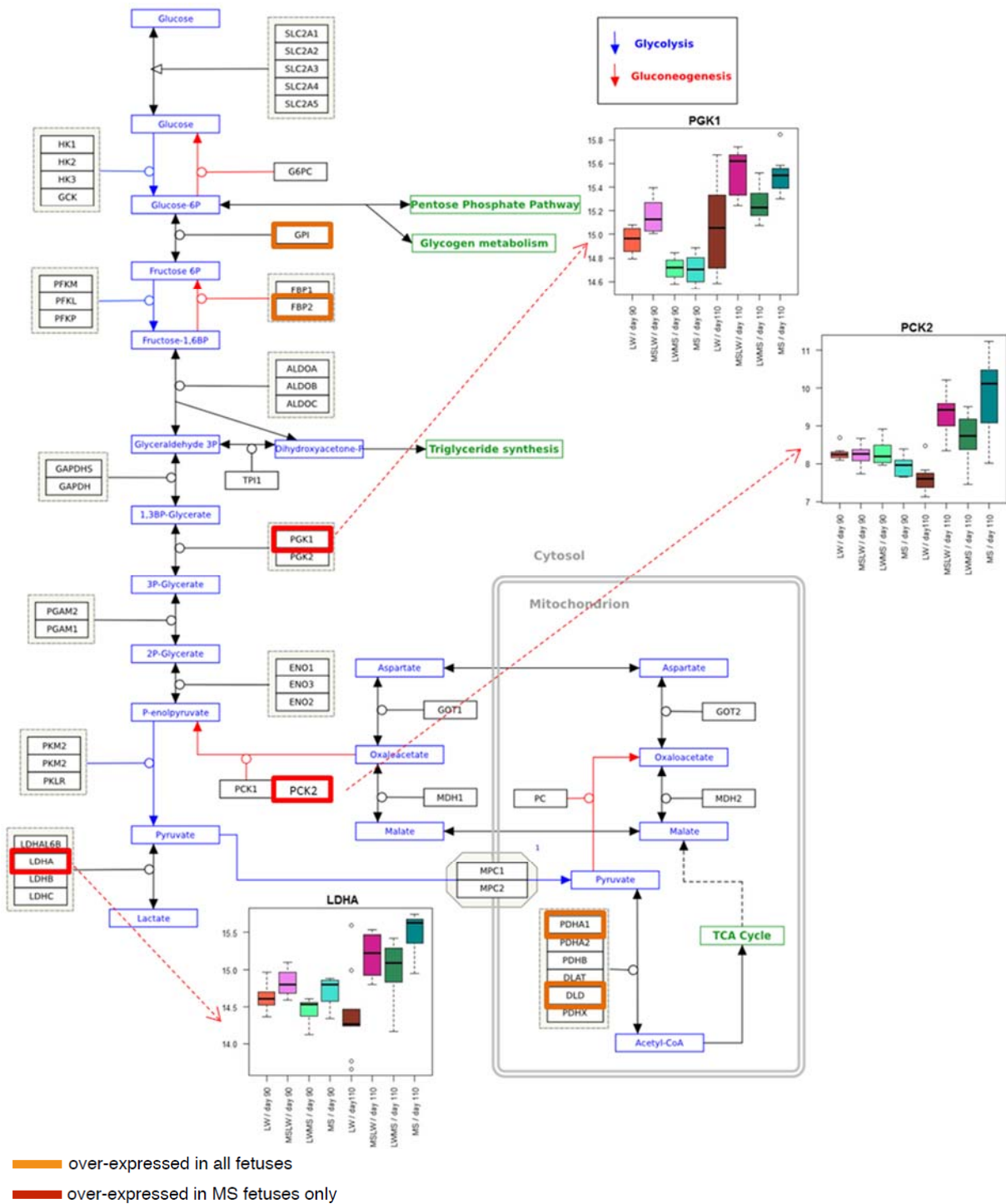


Figure 4 : Illustration des variations observées de la glycolyse et de la gluconéogenèse dans le muscle fœtal à 90 et 110 jdg.

L'analyse fonctionnelle des DEG a été réalisée sur la liste des gènes régulés en fin de gestation et influencés différemment selon le génotype des fœtus. Ce sont les gènes qui nous intéressaient le plus car, a priori, devant nous donner des indications sur ce qui caractérise la meilleure maturité à la naissance observée chez les MS. Nous définissons la maturité comme l'état de plein développement permettant la survie à la naissance. Les gènes différentiels sont identifiés pour l'interaction entre le génotype et le stade de développement. Cette liste comprend 2000 sondes, correspondant à 1090 gènes uniques (annotation décembre 2018, P. Dehais, Sigene).

Globalement, la maturité concerne des gènes du développement surexprimés à 90 jdg et des gènes du métabolisme à 110 jdg. L'étude fonctionnelle a révélé notamment l'importance de la voie de la gluconéogenèse (peu ou pas décrite dans le muscle), ainsi que l'implication de plusieurs fonctions (structure du collagène, division cellulaire, adhésion cellulaire, métabolisme des acides gras (Voillet *et al.* 2014)). Le résultat le plus surprenant est donc celui d'une possible gluconéogenèse dans le muscle alors que ce mécanisme est essentiellement décrit dans le foie.

Les profils d'expression des gènes différentiels impliqués dans la glycolyse et la gluconéogenèse en fin de gestation sont particulièrement intéressants. Dans la Figure 4, les profils des gènes *PGK1* (Phosphoglycerate Kinase 1), *PCK2* (Phosphoenolpyruvate Carboxykinase 2, Mitochondrial ou PEPCK-M) et *LDHA* (Lactate Dehydrogenase A) posent plusieurs questions. Pour *LDHA* et *PCK2*, l'expression dans le muscle de fœtus à 90 jdg n'est pas très différente entre les deux génotypes mais nous observons une augmentation de l'expression chez les MS et les croisés alors que l'expression est peu modifiée chez les LW, voire même diminue pour *PCK2*. L'expression de *PGK1* montre une très grande variabilité intra-LW. Nous pouvons faire le lien avec le défaut de maturité globalement observé chez les porcelets LW à la naissance et l'hétérogénéité de portée, un des éléments clés des problèmes de survie néonatale Canario (Canario *et al.* 2006). Ici, il faut rester prudent puisque cette observation n'est réalisée que sur huit fœtus mais reste une information très intéressante si on la met en relation avec les phénotypes, les données physiologiques et les données métabolomiques acquises sur 400 voire 600 fœtus.

Quoiqu'il en soit, le résultat pour *PCK2* soulève plusieurs questions. Ce gène code une enzyme plus connue sous le nom de PEPCK pour PhosphoEnolPyruvate CarboxyKinase 2, mitochondrial. Florence Gondret et Louis Lefaucheur (PEGASE, St Gilles) expliquent que classiquement, on considère l'activité PEPCK négligeable dans les tissus où il n'y a pas de néoglucogenèse comme le muscle. En fin de gestation, les données musculaires révèlent depuis quelques années l'expression de PEPCK dans le muscle et le tissu adipeux, remettant en cause comme pour d'autres voies, les connaissances de biochimie. Néanmoins, l'activité de PEPCK resterait faible. Contrairement au foie, le muscle est plutôt consommateur que producteur de glucose, ce qui explique que PEPCK impliqué dans la néoglucogenèse soit faiblement exprimé dans le muscle. Quoiqu'il en soit, l'expression au niveau protéique par immunohistochimie a été analysée en particulier chez la souris et détectée dans certains muscles squelettiques suggérant une association avec le type de fibre (Zimmer & Magnuson 1990). Plus précisément, les auteurs observent même un marquage périodique et suggèrent une association préférentielle avec des éléments particuliers de l'appareil contractile. Les auteurs précisent également que cette détection débute à 20 jdg donc en fin de gestation chez la souris.

Ceci dit, le papier de Brown *et al.* (2016) montre la présence de PCK2 dans le muscle aux niveaux ARN et protéique, avec un effet clair de la ractopamine (bêta-agoniste qui permet d'obtenir de la viande plus maigre et plus protéinée, interdit en Europe) sur l'expression de *PCK2*. L'hypothèse de Brown *et al.* sur le rôle possible de PCK2 dans la redirection de l'oxaloacetate mitochondrial vers la synthèse d'acides aminés dans le cytoplasme est intéressante mais il s'agit là d'une hypothèse qu'il reste à valider.

Je ne reprendrais pas tous les résultats de Valentin. Par contre, nous avons commencé à explorer la variabilité intra-LW pour essayer de lister tous les gènes présentant des profils similaires à celui de *PGK1* dans la Figure 4. Ce type de profils est aussi présent dans les résultats du transcriptome des glandes surrénales (Terenina *et al.*, non publié). Malgré le faible dispositif pour prétendre parler de variabilité intra génotype et comme les autres profils d'expression de gènes sont très contrastés entre génotypes et stades, nous restons convaincus que ces gènes peuvent avoir un intérêt particulier pour expliquer le retard de croissance des LW.

Toujours, dans le cadre de sa thèse, Valentin Voillet a également été réalisé une analyse du protéome par électrophorèse 2D sur les mêmes 64 échantillons de muscle (Voillet *et al.* 2018). L'analyse du protéome a été réalisée par Louis Lefaucheur (PEGASE, INRA Centre Bretagne-Normandie, Rennes).

Une particularité de l'analyse protéomique est que l'étude statistique se fait en aveugle pour ensuite sélectionner les spots différentiels qui seront prélevés pour identification par spectrométrie de masse. Initialement, 1025 spots ont été identifiés sur les différents gels.

La stratégie choisie a été de multiplier les analyses statistiques (Voillet *et al.* 2018). Les analyses statistiques réalisées étaient :

- Modèle mixte linéaire : 104 spots (complet, additif, génotype, âge),
- Analyses discriminantes : sPLS-DA (33 spots), RF (16 spots), RF par régression avec les données de typage des myosines (49 spots),
- Analyses multivariées : sPLS avec les myosines (87 spots), avec les 2000 sondes différentielles par analyse du transcriptome (20 spots), avec le glycogène et des données physiologiques (28 spots), et CCA avec les myosines (41 spots).

Ces analyses aboutissent à identifier 113 spots différentiels pour finalement caractériser 89 protéines uniques après analyse par spectrométrie de masse.

L'analyse post-statistique correspond souvent à la démarche suivante (valable à quelques détails près pour les deux publications de Valentin Voillet (Voillet *et al.* 2014; Voillet *et al.* 2018) et de façon générale à chacun de mes travaux.

- Recherche d'enrichissement pour les termes d'ontologie de gènes, les KEGG *pathways* ou voies de signalisation et métaboliques bien connues. Cette recherche d'enrichissement peut être faite sur l'ensemble des gènes ou de sous-groupes.
- Construction de réseaux spécifiques du stade 90 ou 110 jdg. Nous utilisons différentes corrélations pour cela. Pour les données du transcriptome (Voillet *et al.* 2014), Valentin a utilisé une corrélation simple (Pearson). Pour les données du protéome, Valentin a utilisé l'approche PCIT (*partial correlation and information theory*, (Reverter & Chan 2008)) qui calcule une corrélation partielle « allégée » par rapport aux approches GGM (Chapitre 3d). La corrélation entre deux nœuds est calculée en sachant la corrélation avec une troisième protéine. La corrélation est transformée en lien entre deux protéines si elle est significative. Ici un seuil sur la densité a aussi été posé (densité < 5%) pour que le graphe soit lisible.
- Recherche de communautés. Le principe de base est globalement le même. Il s'agit de regrouper les nœuds très connectés entre eux dans une même communauté.
- Analyse fonctionnelle des communautés. L'analyse fonctionnelle est réalisée avec IPA (Ingenuity Pathway Analysis) qui peut proposer des réseaux de co-occurrence bibliographique et aussi peut proposer de possibles régulateurs (*upstream regulators*) qui sont souvent des facteurs de transcription. Des tests d'enrichissement sont associés pour donner un niveau de significativité. La pertinence d'un réseau statistique est donc également évaluée pour leur pertinence biologique. En général, je reconstruis le réseau biologique.

Je ne reprends pas ici les résultats publiés. La démarche est illustrée dans ce document au niveau des Figures 26, 27, 28.

Pour cette étude, nous disposons du dosage du glycogène dans le muscle et du typage précis des fibres musculaires au niveau protéine et transcrit. Les dosages musculaires du glycogène et le typage précis des chaînes lourdes de myosine (8 isoformes de MyHC codées par 8 gènes différents) dans le muscle *longissimus* des fœtus ont été utilisés pour intégrer les données protéomiques. Le glycogène et les chaînes lourdes de myosine évoluent fortement entre 90 et 110 jdg. Une bonne correspondance entre les niveaux protéiques et d'ARN messagers suggère une régulation au niveau transcriptionnel pour cette famille de protéine (Figure 5). Une corrélation positive a également été trouvée entre les transcrits et 31 protéines codées par ces transcrits. L'information que la régulation peut se limiter essentiellement au niveau transcriptionnel peut indiquer une implication plus directe dans la régulation des mécanismes qui nous intéressent.

Ainsi, l'expression de la MyHC embryonnaire diminue fortement entre 90 et 110 jdg alors que celle des MyHC adultes rapides (IIa + IIx + IIb) augmente (Figures 5 et 6), ce qui en font respectivement des marqueurs d'immaturité et de maturité musculaire.

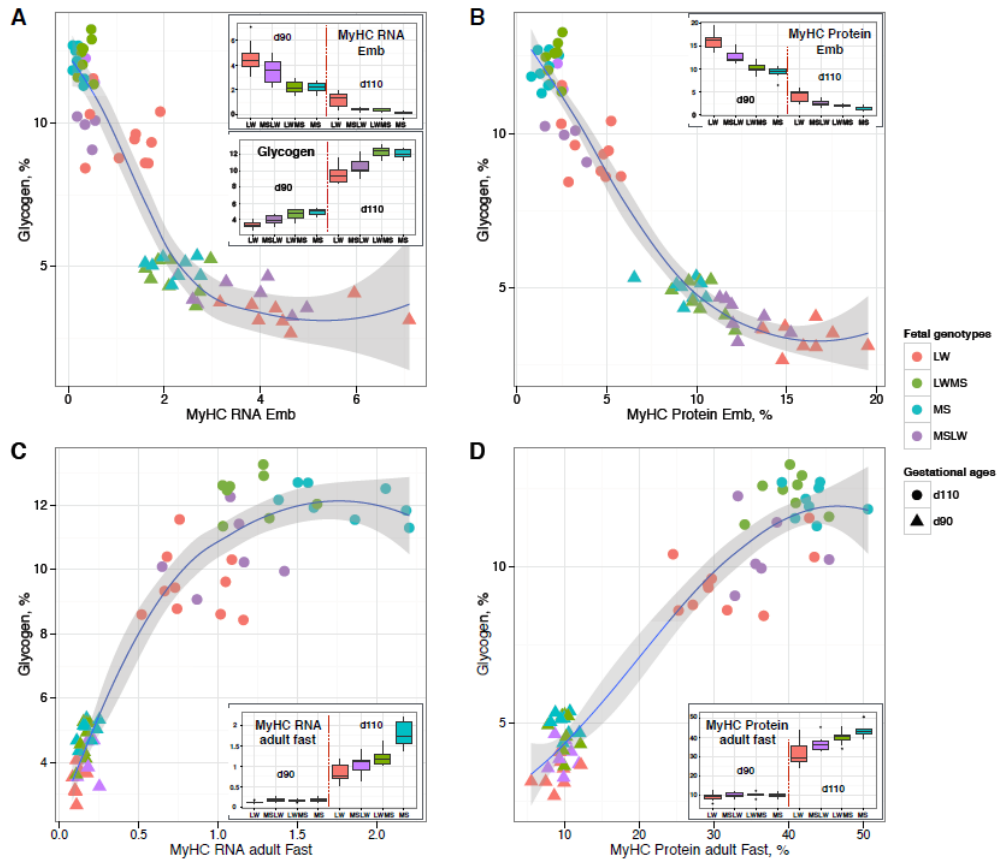


Figure 5 : Coévolution de la myosine embryonnaire, de la myosine adulte (rapide) et du glycogène musculaire en fin de gestation.

Les quatre figures montrent la coévolution du glycogène avec A- la myosine embryonnaire au niveau messager, B- la myosine embryonnaire au niveau protéique, C- la myosine adulte (rapide) au niveau messager et D- la myosine adulte (rapide) au niveau protéique.

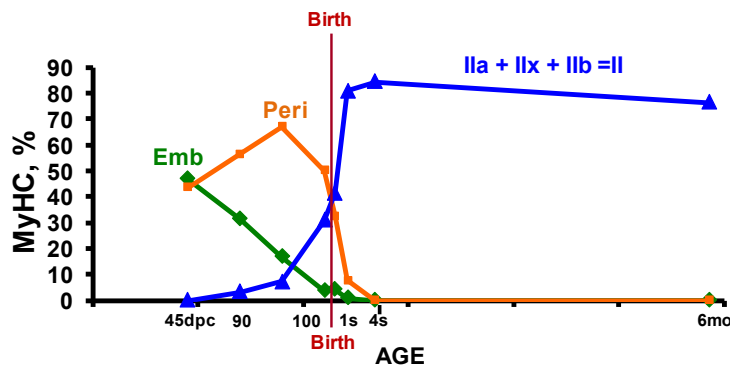


Figure 6 : Evolution des chaînes lourdes de myosines chez le porc.

Les myosines suivent une évolution cordonnée au cours du développement pendant la gestation et en périnatal. Les chaînes lourdes de myosine sont (nom du gène correspondant) : Emb (MYH3), Peri ou fœtale (MYH8), Adultes « fast » : Ila (MYH2), Iix (MYH1) et Iib (MYH4). Compilation des travaux et figure amicalement fournies par Louis Lefaucheur.

Ces marqueurs montrent que les fœtus purs Meishan semblent plus matures que les purs Large White à 90 et 110 jdg et que les croisés sont le plus souvent intermédiaires (Figure 5). Une bonne corrélation entre ces critères de maturité MyHC et la teneur du glycogène musculaire suggère une régulation fine des mécanismes en fin de gestation. L'analyse fonctionnelle combinée des résultats protéomiques avec les données transcriptomiques confirme le retard de développement déjà suggéré avec les données du

transcriptome. La Figure 7 montre l'analyse combinée des données transcriptome et protéome par PLS associée à une analyse d'enrichissement d'annotations en ontologies de gènes. Nous retrouvons l'information que les LW seraient toujours plutôt en phase de développement musculaire (à gauche sur la Figure 7A et 7B) quand les MS mettraient en fonctionnement les mécanismes énergétiques juste avant la naissance (à droite sur la Figure 7A et 7B).

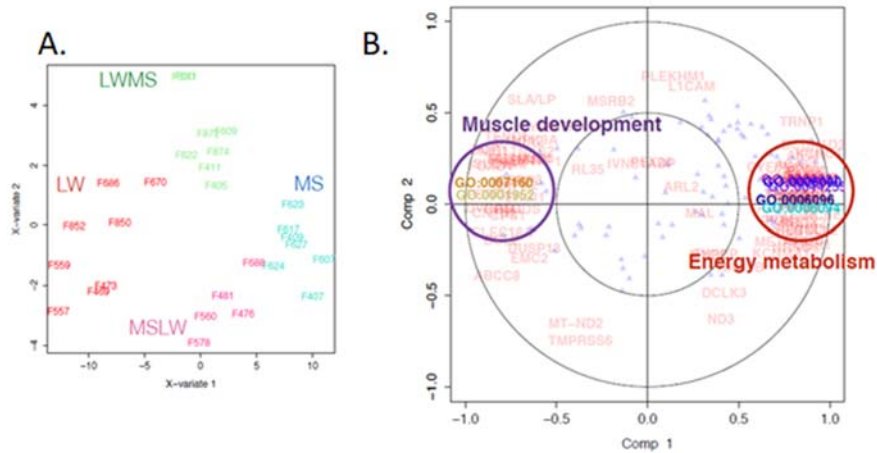


Figure 7 : Analyse PLS entre les données musculaires du transcriptome et du protéome à 110 jdg. Cette analyse a été réalisée avec 90 gènes and 86 protéines surexprimés à 110 jdg. A. La projection des fœtus montre la séparation des génotypes. B. L'enrichissement fonctionnel est superposé à la projection des gènes et des protéines.

Finalement, Valentin Voillet a tenté une approche graphique pour intégrer l'ensemble des données disponibles pour le muscle, données du transcriptome, du protéome et des phénotypes. La démarche s'est inspirée des travaux de Montastier et al. (2015) où d'abord est calculé les corrélations partielles au sein de chaque jeu de données (voir le chapitre 3d pour plus d'explication sur l'inférence de graphes avec une approche GGM, *Gaussian Graphical Model*) puis la combinaison des différents niveaux est réalisée avec une approche CCA (*Canonical Correlation Analysis*) pour réaliser un réseau bipartite et ainsi créer les liens entre les différents jeux de données (Figure 8).

Le réseau de la Figure 8 met en avant l'importance (forte centralité) du glycogène et du fructose aussi ; le fructose d'origine placentaire participe probablement à la synthèse du glycogène chez le fœtus (Battaglia & Meschia 1978). Le fructose diminue entre 90 et 110 jdg quand le glycogène augmente. Ces deux molécules semblent donc avoir un rôle central dans la structuration du réseau qui est facilement compréhensible par rapport au rôle que joue le glycogène, synthèse et stockage dans le muscle en fin de gestation. Le glycogène est la première réserve énergétique utilisée à la naissance pour assurer les premiers besoins du nouveau-né (thermorégulation...).

Néanmoins nous pouvons observer que globalement les données protéomiques sont séparées des autres données (en rouge en haut du réseau). Quelques petits groupes de protéines sont localisés avec des transcrits ; nous pouvons même visualiser que CKMT2 gène et protéine sont liés. La protéine CKM2 a déjà été identifiée par Valentin Voillet pour son rôle central dans les réseaux protéiques à 110 jdg (sous-réseau 2, (Voillet *et al.* 2018)). De plus, CKMT2 est effectivement une des protéines dont l'abondance est retrouvée corrélée avec l'expression du gène CKMT2, suggérant une régulation de l'expression surtout au niveau transcriptionnel.

Les analyses du transcriptome, du protéome et des phénotypes ont mis en évidence plusieurs voies métaboliques qui doivent se mettre en place en fin de gestation pour assurer la survie néonatale. Certains gènes et/ou protéines semblent avoir un rôle plus important dans les différents processus cellulaires. Et surtout l'ensemble des résultats pose la question de l'hétérogénéité de l'expression de certaines variables entre les deux lignées extrêmes pour la survie néonatale. Des résultats posent questions pour des mécanismes qui suggèrent un retard au développement juste avant la naissance, ils montrent parfois une

grande variabilité intra-LW.

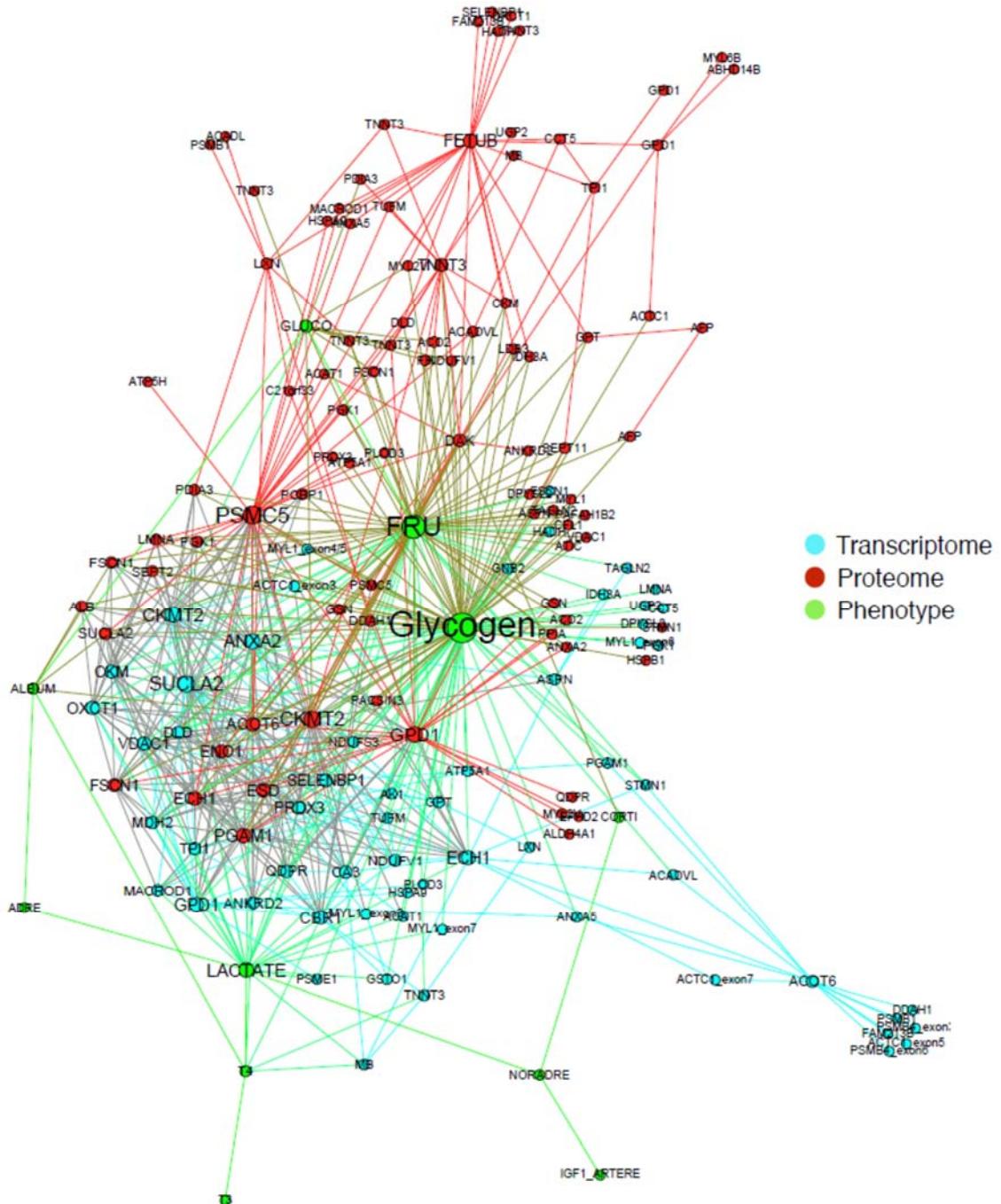


Figure 8 : Réseau à 110 jdg prenant en compte les relations intra-jeu de données transcriptome, protéome et les phénotypes (réseaux de co-expression) et entre les différents niveaux d'information (réseaux bipartites). La démarche s'est inspirée du modèle développé dans Montastier et al. (2015). Ce réseau est un résultat non publié de la thèse de Valentin Voillet. Le code couleur illustre les liens entre les variables d'un même jeu de données alors que les liens entre différents types de variables sont en vert kaki entre protéine et phénotype, vert-bleu entre phénotype et transcrit et gris entre transcrit et protéine. La taille des nœuds illustre l'importance de la variable dans la structure du réseau.

Les analyses du transcriptome et du protéome ont été limitées à une soixantaine de fœtus. L'intégration de ces données pourra se faire avec l'ensemble des informations disponibles dans le projet PORCINET (dosages plasmatiques ou tissulaires, mesures morphologiques, métabolome plasmatique) qui proviennent de 400 à 600 fœtus. Ces différents niveaux d'intégration permettront de replacer l'information moléculaire fine pour un tissu dans un contexte plus général, en particulier les différentes variations citées plus haut.

c. Analyse du transcriptome musculaire en fin de gestation chez trois espèces d'intérêt agronomique - Diversité intra espèce vs inter-espèces

Toujours dans le cadre de la thèse de Valentin, nous nous sommes intéressés à repositionner les résultats obtenus pour le muscle fœtal chez le porc chez d'autres espèces d'intérêt agronomique. Ce travail a été réalisé en collaboration avec Brian Dalrymple chez qui Valentin Voillet est allé faire un séjour de 4 mois à Brisbane en Australie (CSIRO Animal, Food and Health Sciences).

La figure 9 illustre les évènements communs aux trois espèces pendant la gestation et surtout la période qui nous intéresse, la fin de gestation.

Pour réaliser cette étude comparative, nous avons récupéré des données d'analyse du transcriptome de muscle fœtal bovin (Lehnert et al. 2007). Dans cette étude, les femelles Hereford avaient été croisées soit avec des mâles Wagyu (lignée avec plus de gras intramusculaire) soit des mâles de la race Piémontaise qui présente une hypertrophie musculaire due à la présence d'une mutation dans le gène *GDF8* (Growth differentiation factor 8 ou myostatine). Nous avons aussi récupéré des données d'expression chez le mouton (Ren et al. 2011). Il s'agit de données obtenues à partir de muscle fœtal prélevé sur des moutons de race Texel qui est aussi porteuse d'une mutation dans le gène *GDF8* (Clop et al. 2006) donnant des animaux plus musclés et moins gras. La race Ujumqin, d'origine mongolienne, plus rustique a également été utilisée. Pour le porc, il s'agissait d'utiliser les données LW et MS publiées (Voillet *et al.* 2014) et discutées dans le chapitre précédent.

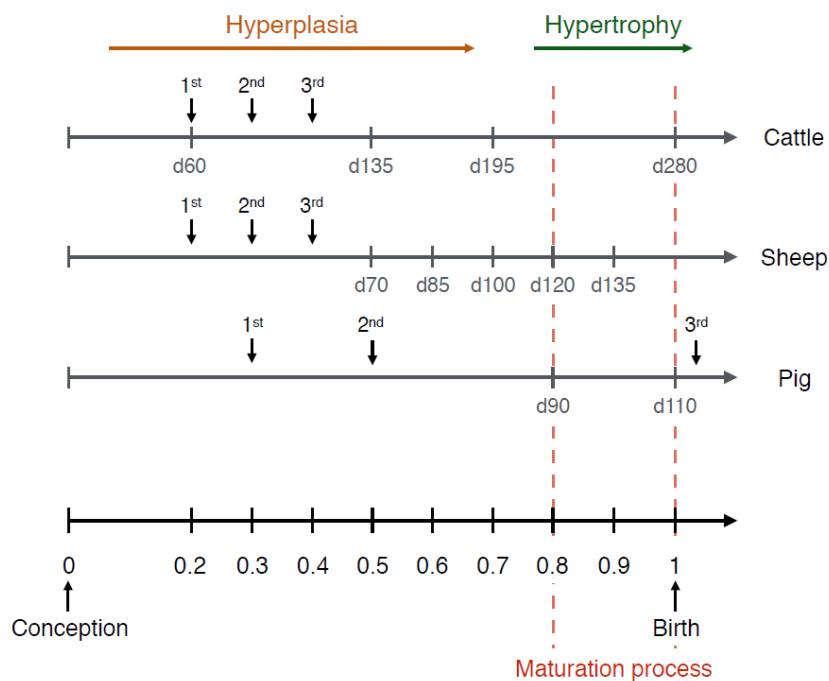


Figure 9 : Vue d'ensemble de la gestation chez trois espèces (bovin, mouton et porc).

Une échelle entre 0 et 1 avec 1 correspondant à la naissance a été utilisée. Les données d'expression génique utilisées dans les jeux de données analysés sont indiquées avec les pointillés rouges. Le moment de la myogenèse primaire (1ère), secondaire (2ème) et tertiaire (3ème) est indiqué pour chaque espèce (Thèse de Valentin Voillet, publication en préparation).

La question était de savoir si des différences de maturité toujours en lien avec la survie néonatale pouvaient être expliquées par des différences entre espèces ou entre lignées génétiques au sein d'une même espèce.

Chez ces trois espèces, il est reconnu que la forte sélection en vue d'une augmentation du nombre d'animaux produits par femelle, d'une masse musculaire accrue et d'une réduction de la teneur en gras

des carcasses aurait une incidence plus ou moins importante sur la survie néonatale. L'objectif de notre étude comparative est d'utiliser les trois dispositifs pour identifier les processus biologiques et les gènes importants pour la maturité à la naissance. Les données retenues sont celles de fin de gestation entre 210 jours pour le bovin, 120 jours pour le mouton et 90 jours pour le porc et la naissance ou le temps le plus proche de la naissance. Si nous pouvons voir sur la figure 9 que ces temps ne sont pas complètement les mêmes, rapportées à une échelle 0-100% de gestation, la période visée est bien celle correspondant à 80% de la gestation où le nombre de fibres musculaires est fixé et la masse musculaire augmente essentiellement par hypertrophie des fibres existantes (Picard et al. 2002).

Gene Set	Biological process	Genes
Cell cycle ¹	Cell cycle process	<i>CDC6, CDC20, CDCA3, CCNB1, CCNB2, KIF20A, CKAP2</i>
ECM ¹	ECM organization	<i>BGN, COL5A2, TGFB2, SERPINH1, SDC3, SH3PXD2B</i>
Muscle structure fast subunits ²	Fast twitch	<i>TMOD4, MYOZ1, MYH1, CKM, ENO3</i>
Mitochondrial transport ¹	Respiratory electron transport chain	<i>ATP5B, PPARGC1A, COX5B, UCP3, NDUFV2</i>
Lipid metabolism ¹	Cellular lipid metabolic process	<i>ACADS, CRAT, SLC25A20, CPT1B, MCEE</i>
Glycogen metabolism ¹	Glycogen metabolic process	<i>GAA, PYGM, PPP1R1A, PPP1R3A, GSK3A</i>

Table 1 : Six sets de gènes comparés entre et intra espèces.

Ces ensembles de gènes ont été identifiés à partir d'un même terme GO¹ ou bien d'après les travaux² de (Hudson et al. 2013).

A partir des trois jeux de données, un certain nombre de petits ensembles de gènes considérés comme robustes a été retenu pour refléter les processus biologiques auxquels ces gènes participent (Hudson et al. 2013). Les gènes sélectionnés sont présentés dans la Table 1.

La particularité des espèces ovines et bovines est qu'elles ont été sélectionnées pour une carcasse plus musclée en profitant d'une mutation dans le gène de la myostatine (ou *GDF8*). La myostatine est une protéine de la famille des TGFbeta (*Transforming Growth Factor*) qui a un rôle de régulateur négatif de la masse musculaire. Des mutations dans le gène *GDF8* provoquent une hypertrophie musculaire en inhibant la prolifération et la différenciation des myoblastes (Clop et al. 2006 ; Kambadur et al. 1997). La mutation chez les bovins induit une augmentation du poids des veaux à la naissance rendant la mise-bas difficile avec augmentation de la mortalité néonatale.

Sans grande surprise, pour la période de mise en place de la maturité, le processus de développement en fin de gestation est très conservé entre les espèces. On notera dans la Figure 10 l'expression décroissante des gènes du cycle cellulaire (prolifération) et de la matrice extracellulaire (support du développement musculaire (Frantz et al. 2010; Rozario & DeSimone 2010)) et l'expression croissante des gènes de différentes voies métaboliques (des lipides, du glycogène et mitochondriale) ainsi que les protéines des fibres rapides (fibres adultes).

Par contre, il y a des différences significatives intra espèces, entre les génotypes.

Pour la comparaison des bovins dérivés de la race Piémontaise hétérozygote pour la mutation *GDF8* ou dérivés de la race Wagyu à viande plus persillée, des différences significatives sont identifiées pour le cycle cellulaire, la matrice extra-cellulaire, et le métabolisme des lipides à la naissance.

Chez le mouton, la comparaison entre les Texel (porteurs d'une mutation dans *GDF8*) et les Ujumqin (plus rustiques) donne aussi des différences significatives pour le cycle cellulaire et la structure des fibres rapides.

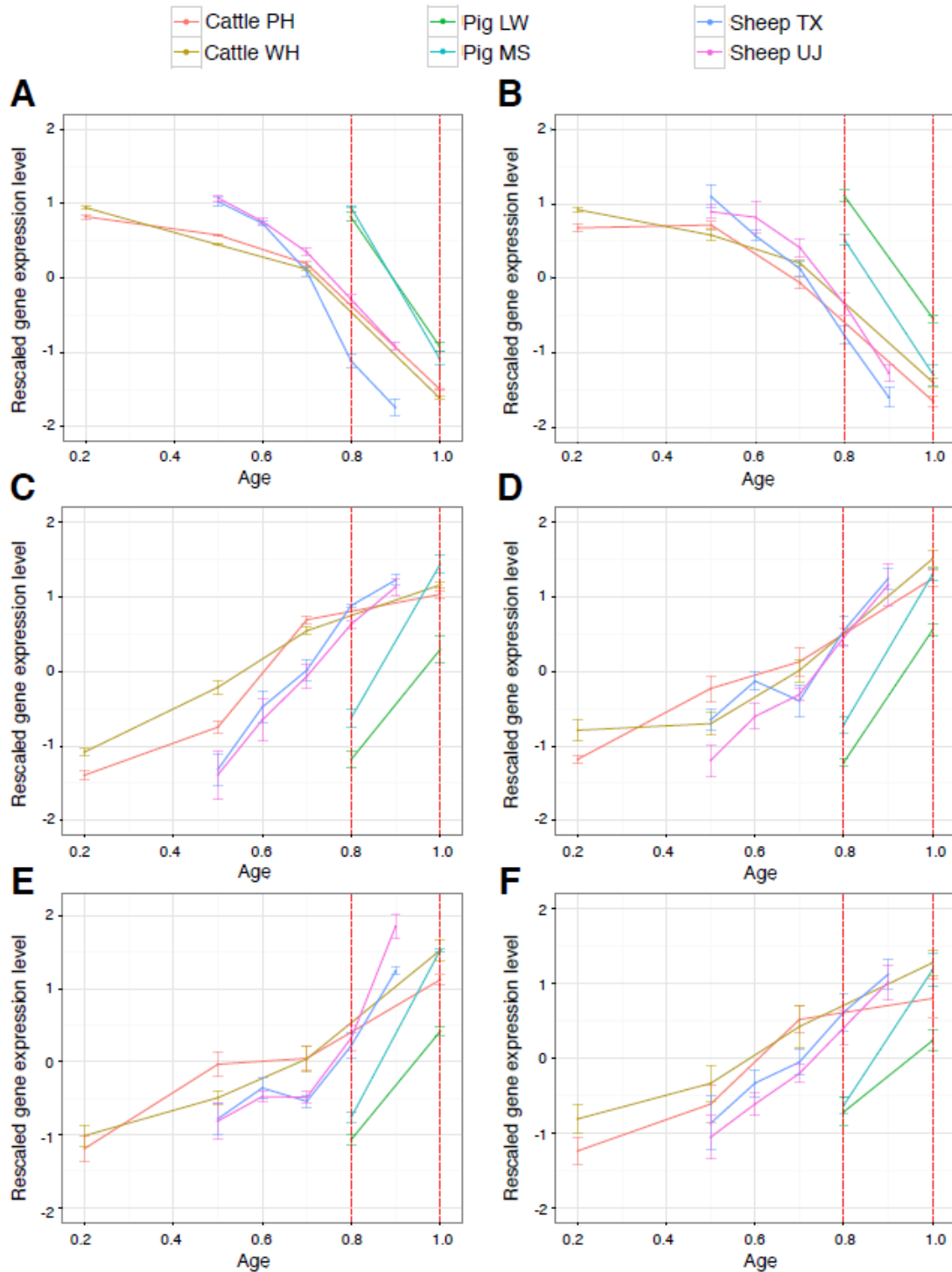


Figure 10 : Profils d'expression des six sets de gènes représentant (A) le cycle cellulaire, (B) la matrice extracellulaire, (C) les protéines des fibres rapides, (D) le transport mitochondrial, (E) le métabolisme lipidique et (F) le métabolisme du glycogène.

Voir la Table 1 pour les gènes retenus pour faire ces analyses. Les niveaux d'expression des sets de gènes sont les scores z moyens des gènes de chaque ensemble de gènes.

L'utilisation de groupes de quelques gènes co-exprimés pour représenter spécifiquement différents processus biologiques semble une bonne idée. Cette démarche (Guo et al. 2015) semble permettre de diminuer le bruit qui peut se trouver dans les grands jeux de données en se focalisant sur des gènes les plus représentatifs d'un groupe de gènes, d'identifier des processus spécifiques de type cellulaire alors que l'analyse du transcriptome interroge un tissu hétérogène (comme le muscle)... **Il pourra être**

intéressant de suivre une démarche à peu près similaire pour simplifier nos grands jeux de données en se basant sur des groupes de gènes présentant une très forte corrélation. Je ne suis pas sûre qu'il soit complètement adéquat de se baser que sur des gènes partageant une même fonction biologique pour définir des groupes. Quoiqu'il en soit, cette démarche peut être séduisante pour les futures étapes d'intégration de données.

d. Architecture nucléaire et régulation de l'expression des gènes musculaires en fin de gestation

Ici je vais aborder un aspect de l'étude du muscle qui est une base pour de futures recherches et qui aurait pu être incluse dans les perspectives. Toutefois, une première étude a été publiée et a fait l'objet du sujet de stage de Master et de thèse de Maria Marti-Marimon avec qui j'ai eu le plaisir de collaborer. Son travail a été réalisé dans l'équipe CYTOGENE de notre laboratoire sous la responsabilité d'Yvette Lahbib-Mansais ainsi que de Martine Yerle-Bouissou pour les aspects cytogénétiques et Nathalie Vialaneix (Unité MIA-T) pour le modèle d'inférence de graphes. Valentin Voillet avait réalisé les premiers réseaux de gènes pour ce projet.

Les échantillons de muscle utilisés sont ceux du projet PORCINET donc des échantillons prélevés à 90 et à 110 jdg et pré-traités pour des études de localisation de gènes dans l'espace nucléaire. Les données d'expression sont celles de Valentin Voillet. La question posée était très ambitieuse puisqu'il s'agissait de partir de l'hypothèse que des gènes co-exprimés étaient co-régulés donc à priori par une même machinerie transcriptionnelle et donc que les régions génomiques portant les gènes co-exprimés devaient être co-localisées dans l'espace nucléaire. L'idée était donc d'aller effectivement vérifier qu'il en était ainsi. Mais nous avons conscience que la question était assez délicate puisque le transcriptome musculaire est réalisé sur un mélange de cellules et de types cellulaires différents alors que les analyses cytogénétiques sont réalisées sur des cellules individualisées, des fibres musculaires plurinucléées dans notre cas. La technique s'appelle la FISH 3D pour la *3D fluorescent in situ hybridization*. L'étude cytogénétique se fait de noyau à noyau avec le calcul d'un pourcentage de proximité observée entre deux fragments génomiques marqués. Nous avons appelé le projet INTERNET3D et il a été financé par le département de Génétique Animale de l'INRA.

Côté cytogénétique, une première étude s'était portée sur la co-localisation des allèles exprimés de gènes soumis à empreinte (Lahbib-Mansais et al. 2016). Un premier gène central dans la régulation de la croissance musculaire est *IGF2* (*insulin like growth factor 2*). En génétique porcine, nous nous intéressons beaucoup au gène *IGF2*. *IGF2* est un gène d'expression paternelle qui influence la croissance, le développement musculaire et du tissu adipeux ; une mutation dans ce gène a été caractérisé pour expliquer les effets observés à un QTL sur le chromosome 2 (croissance musculaire et niveau d'adiposité ; Jeon et al. 1999; Nezer et al. 1999; Van Laere et al. 2003).

Dans la première étude, il avait déjà été montré que l'allèle exprimé *d'IGF2* s'associe avec les allèles exprimés d'autres gènes soumis à empreinte comme *DLK1/MEG3* (*delta like non-canonical Notch ligand 1 et maternally expressed 3* (Lahbib-Mansais et al. 2016)). Les gènes soumis à empreinte présentent donc l'avantage de permettre la visualisation des deux allèles séparément.

Côté expression génique, nous nous sommes également focalisés sur *IGF2*. Un réseau de gènes a donc été construit en recherchant les gènes co-exprimés avec *IGF2* (Voillet et al. 2014). Le processus itératif de création des graphes est présenté dans la figure 11.

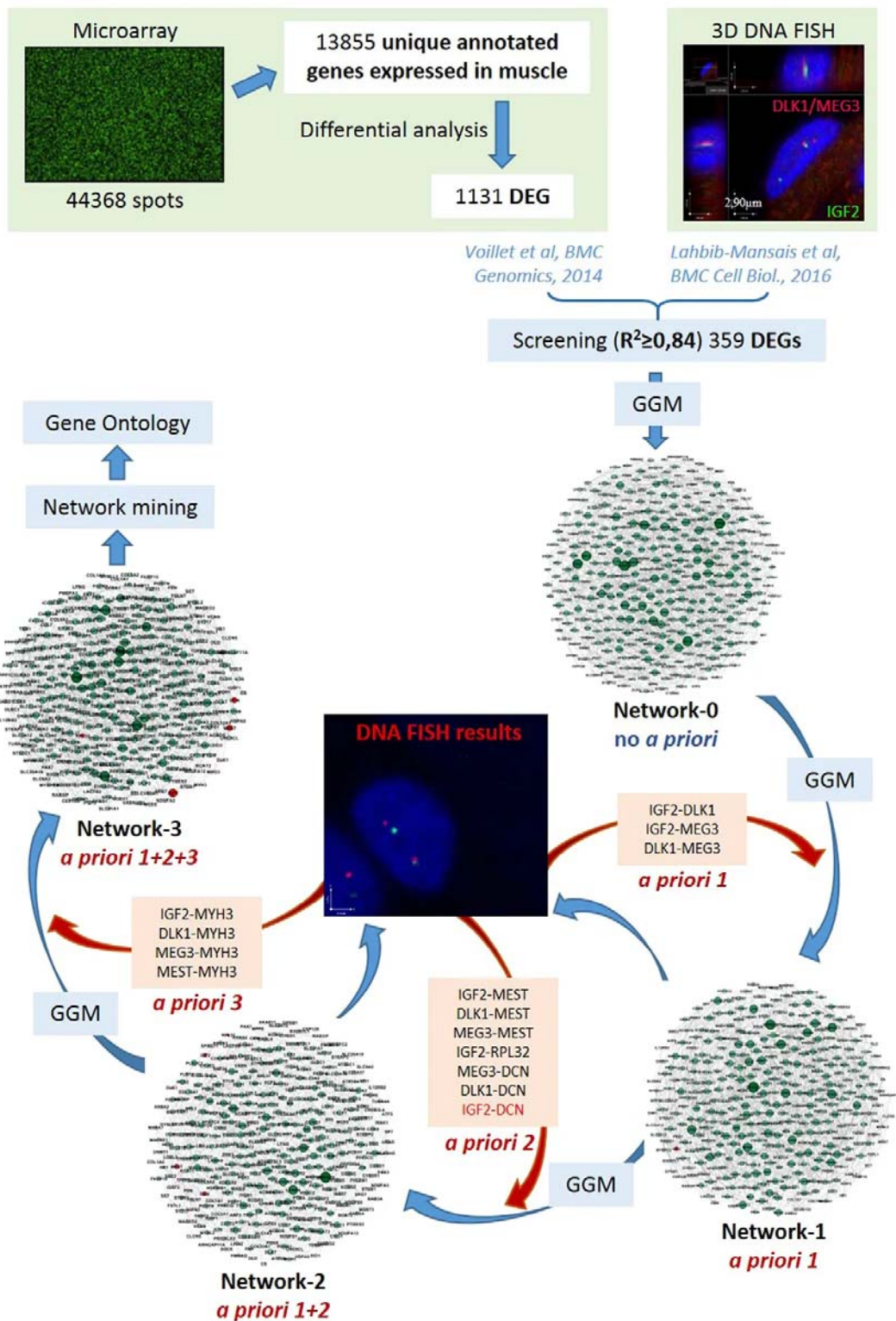


Figure 11 : Présentation du processus itératif d'inférence des graphes de co-expression autour d'IGF2. Les données initiales publiées sont représentées dans des carrés verts (données du transcriptome et données de 3D DNA FISH obtenues dans le muscle fœtal), les méthodes statistiques sont représentées en bleu (GGM : Modèles Graphiques Gaussiens) et les nouvelles informations sur la localisation spatiale utilisée pour l'inférence de réseau sont représentées en rouge.

Le jeu de données initial s'est concentré sur 359 gènes dont l'expression est corrélée (corrélation de Pearson) à *IGF2* ainsi qu'à *DLK1* et *MEG3*. Un premier graphe est construit selon un modèle classique de corrélation partielle (corrélacion entre deux entités connaissant les corrélacions avec toutes les autres

entités). L'approche utilisée est basée sur les modèles graphiques gaussiens (*Gaussian Graphical Model*). On obtient un premier réseau. Parmi les gènes connectés à *IGF2* se trouve *RPL32* (*ribosomal protein L32*) qui fait partie des gènes du réseau de Varrault et al. (2006), réseau de gènes co-exprimés soumis à empreinte. *RPL32* a donc été choisi pour être testé par FISH. Deux autres gènes (*MEST* et *DCN*) ont également été testés en FISH car ils sont également dans le réseau de Varrault et dans notre liste des 359 gènes. Les trois associations se sont avérées positives et ont donc été utilisées pour construire le graphe suivant en mettant un *a priori* sur ces associations dans le modèle. On obtient un nouveau graphe et ainsi de suite jusqu'au troisième réseau.

Une des informations originales sortie de ce travail est une association entre *IGF2* et *MYH3* (*myosin heavy chain 3*), la myosine embryonnaire dont l'importance a déjà bien été caractérisée dans notre projet par Valentin Voillet et Louis Lefaucheur et déjà présentée dans le chapitre précédent (Voillet *et al.* 2018) (Figure 12).

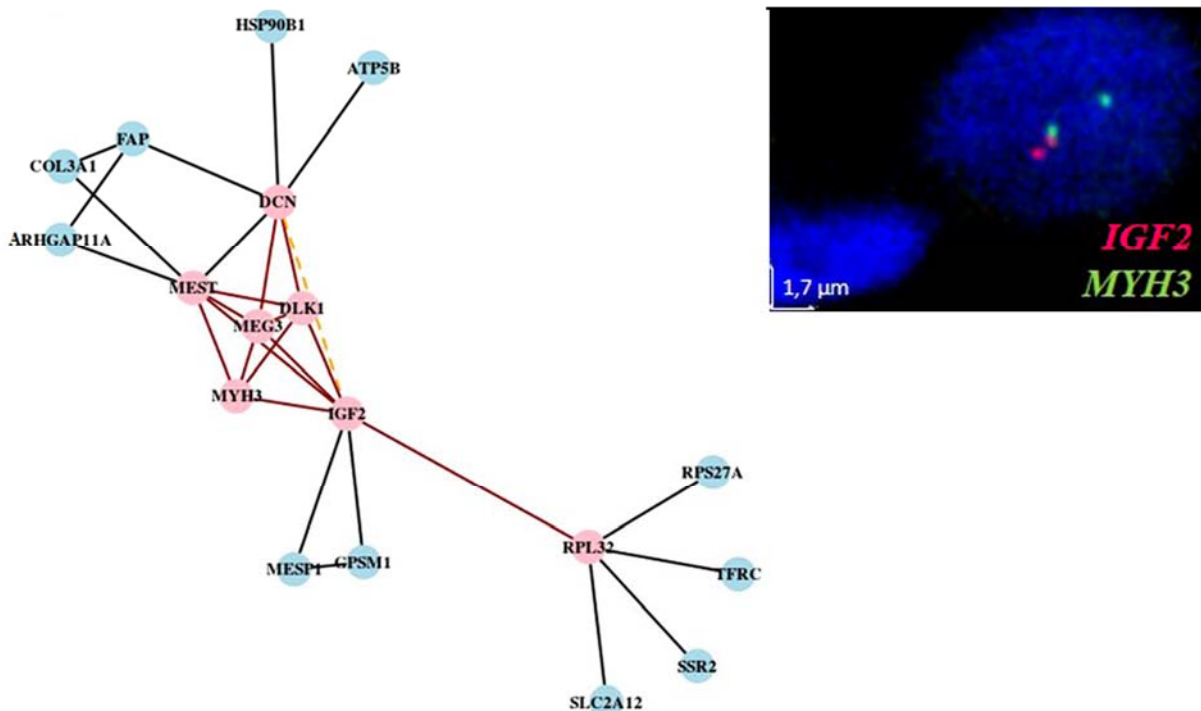


Figure 12 : Sous-réseau 3 du réseau de co-expression et co-localisation nucléaire des allèles exprimés. A gauche ne sont présentés que les liens directs avec nos gènes cibles (en rose) dont *IGF2*. A droite, la co-localisation nucléaire par FISH 3D d'un allèle d'*IGF2* avec *MYH3*.

Une recherche de communautés (clusters) a été réalisée sur tous les réseaux de 0 à 3 afin d'évaluer l'impact du modèle sur le regroupement des gènes. Les gènes cibles se retrouvent tous dans le cluster 1 du réseau 3 avec donc *IGF2*, *DLK1*, *MEG3* et *MYH3*. D'un point de vue fonctionnel, on observe un enrichissement plus significatif de certains processus biologiques dans ce cluster. En particulier, ce cluster est particulièrement enrichi en gènes codant des collagènes, protéines de la matrice extracellulaire ($FDR < 1,14 \cdot 10^{-08}$). Ce qui indique un rôle important de la matrice extracellulaire dans la phase du développement musculaire à 90 jours où tous ces gènes sont surexprimés par rapport à 110 jours. La matrice extracellulaire avait déjà été mise en avant par Voillet et al. (2014) où dans la figure 4 du papier, la communauté 3 inclut des gènes de la matrice extracellulaire. Par contre les gènes *IGF2*, *MYH3*, *DLK1* et *MEG3* n'y sont pas. L'analyse du cluster 1 avec Ingenuity permet d'apporter des informations complémentaires. En particulier, IPA permet de proposer le seul lien que nous ayons trouvé entre *IGF2* et *MYH3* qui est dans l'analyse dite *upstream regulator* et qui propose deux gènes, *MYOD1* et *CTNNB1*, comme de possibles régulateurs des deux gènes. *MYOD1* (Myogenic Differentiation 1) est connu pour réguler la différenciation des cellules musculaires en induisant un arrêt du cycle cellulaire, condition préalable à l'initiation myogénique. La protéine codée par *CTNNB1* (Catenin Beta 1) fait partie d'un complexe de protéines qui constituent des jonctions adhérentes qui régulent la croissance

cellulaire et l'adhérence entre cellules. CTNNB1 a un rôle dans l'adhésion cellulaire, la signalisation cellulaire et la transcription des gènes ; elle peut être transloquée dans le noyau où elle active la transcription de gènes cibles. Shang et al. (2007) ont révélé que dans les cellules stromales mésenchymateuses de rats, une expression ectopique de *Ctnnb1* inhibe la différenciation adipogénétique et induit la formation de cellules multinucléées exprimant des gènes myogéniques, tels que *MyoD* et *Myhc*, donc en favorisant l'expression de facteurs de transcription spécifiques du muscle.

Ce travail s'est donc positionné au carrefour d'approches très différentes pour explorer une question biologique. Le risque était grand de ne pas retrouver à l'échelle nucléaire ce que nous observions dans la dimension du transcriptome. Nous avons démontré qu'en combinant des approches très différentes, il était possible d'identifier des interactions alors inconnues et donc de proposer de nouvelles cibles à explorer pour comprendre le développement musculaire. Ceci a été possible en ayant accès à un développement méthodologique propre à notre question. Les réseaux de co-expression sont donc des outils présentant un grand intérêt.

Ces travaux ont été poursuivis par Maria Marti-Marimon au cours de sa thèse par la réalisation d'une expérience HiC (capture de conformation chromosomique sur l'ensemble du génome) pour explorer à l'échelle du génome les possibles interactions génomiques. Cette deuxième partie du travail s'est réalisée sous la direction de Sylvain Foissac (équipe Dynagen de notre Unité), toujours en collaboration avec Nathalie Vialaneix pour les approches statistiques. Je ne suis pour l'instant pas du tout intervenue dans cette deuxième partie du travail de la thèse de Maria si ce n'est de participer à quelques discussions sur l'avancée des travaux. J'espère pouvoir définir une stratégie d'intégration de ce nouveau type de données. Un peu avec la même idée que les approches eQTL qui permettent de définir des régions du génome intervenant dans la régulation d'un ou plusieurs gènes en *cis* ou en *trans* (voir chapitre 3d), les données HiC permettront d'identifier des régions *cis* très interconnectées et des régions en *trans* pouvant refléter la possibilité de régions génomiques très éloignées voire sur des chromosomes différents de se rapprocher dans l'espace nucléaire. **Ce sont les interactions en *trans* qui m'intéressent le plus, permettant de rapprocher et donc pourrait expliquer que des machineries transcriptionnelles transcrivent certains gènes ou groupes de gènes selon le territoire chromosomique où se trouvent les séquences génomiques ou pas. Ce niveau d'information est à mettre en relation avec l'exploration des caractères complexes dont le déterminisme génétique est plus difficile à caractériser.** Dans le cas de notre étude, nous avons vu qu'il y avait un changement vraiment important dans le nombre de gènes différentiellement exprimés entre 90 et 110 jdg. Ce qui fait échos à une revue récente sur l'HiC où les auteurs rappellent que les questions relatives au développement des organismes demandent un niveau très strict de régulation spatio-temporelle se caractérisant par des modifications spectaculaires dans l'organisation du génome (Sivakumar et al. 2019). Si un changement majeur devait avoir lieu dans l'organisation nucléaire en lien avec la régulation des gènes exprimés, la comparaison entre les deux stades de gestation avec une approche HiC semble la plus opportune. Les données ont été générées et sont toujours en cours d'analyse, les premiers résultats sont prometteurs.

2. Diversité génétique fonctionnelle

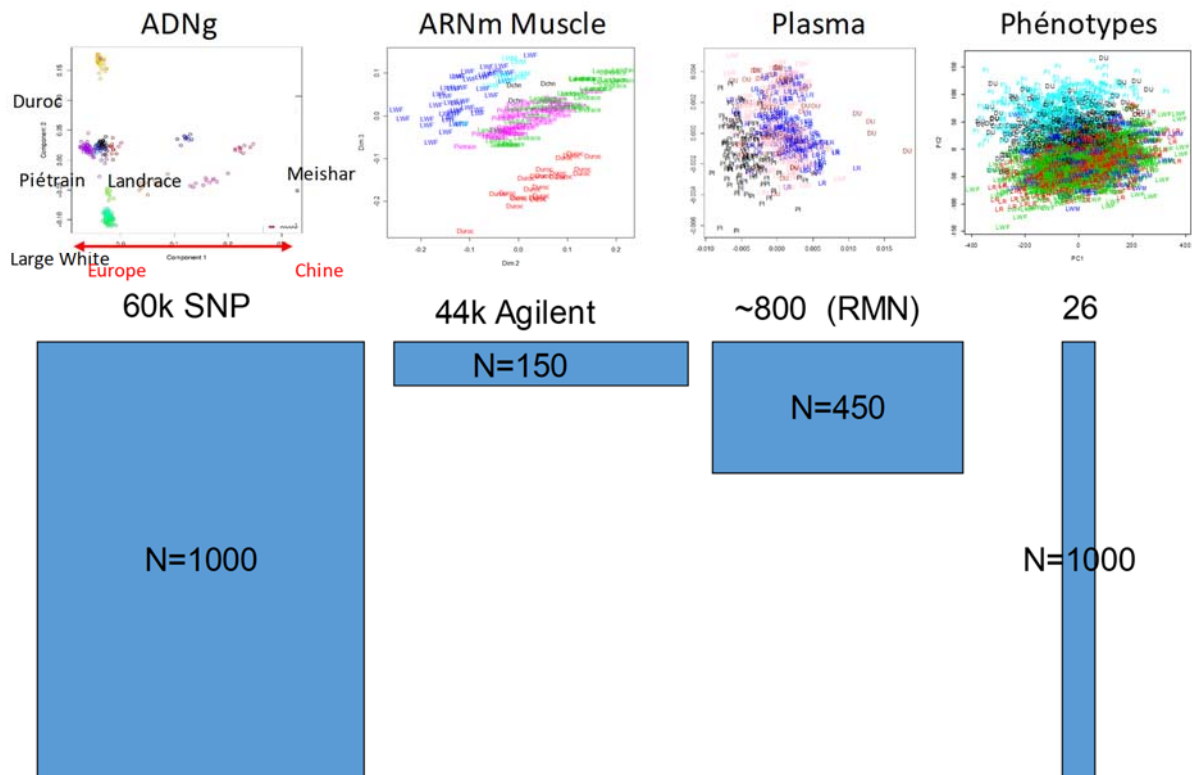


Figure 13 : Analyse de la diversité porcine exprimée à différents niveaux d'expression du génome. Analyse en composantes principales selon le type d'information associé. Les phénotypes « fins » intermédiaires et haut débit permettent de rapprocher l'information phénotypique de ce qui est observé avec le génotypage. L'information donnée par les métabolites se rapproche plus de l'information phénotypique.

Dans le cadre du projet ANR DeLiSus (2008-2010), dont les coordinateurs ont été Denis Milan puis Magali San Cristobal, l'objectif était l'« Etude de la variabilité haplotypique du génome porcin à haute densité ». Ce projet permettait de mener une stratégie de génotypage et de phénotypage à haut débit (il y a maintenant 10 ans...) donc de constituer une base de données de relativement grande taille avec génotypes, données transcriptomiques, métabolomiques et phénotypiques. Les objectifs étaient de détecter des QTL pour les phénotypes, de détecter des traces de sélection et d'étudier l'impact de la diversité génétique sur l'expression des gènes. Les animaux disponibles pour les analyses du transcriptome sont 24 Duroc, 33 Landrace, 41 Large White lignée femelle, 10 Large White lignée mâle and 39 Piétrain.

La figure 13 résume les résultats par des ACP (Analyse en composantes principales) pour chacun des jeux de données et surtout montre l'évolution de la projection des individus selon le type de variables utilisées. Cette Figure permet d'illustrer le potentiel des -omiques intermédiaires à explorer la *black box* existant en génétique, entre l'information génomique et l'expression des phénotypes terminaux. Nous pouvons avoir une idée de la façon dont la complexité des caractères se met en place. Il faut aussi faire attention au caractère déséquilibré des données ainsi acquises qui correspond à la dénomination d'analyse de données hétérogènes. Certains phénotypes « terminaux » peuvent être mesurés sur les mêmes animaux pour lesquels le génotypage a été obtenu. Par contre, l'analyse des phénotypes intermédiaires, transcriptome et métabolome, ne peut être réalisée que sur un nombre plus réduit d'animaux (ce n'est plus tout à fait vrai aujourd'hui pour les données métabolomiques).

Ces données ont été utilisées pour explorer le potentiel de prédiction des données moléculaires intermédiaires. Le transcriptome a été utilisée pour prédire le génotype de 10 individus pour lequel nous n'avions pas l'information de l'origine génétique de l'individu (étude non publiée), le métabolome a été

utilisé pour évaluer la capacité de prédiction de certains phénotypes terminaux (Rohart et al. 2012).

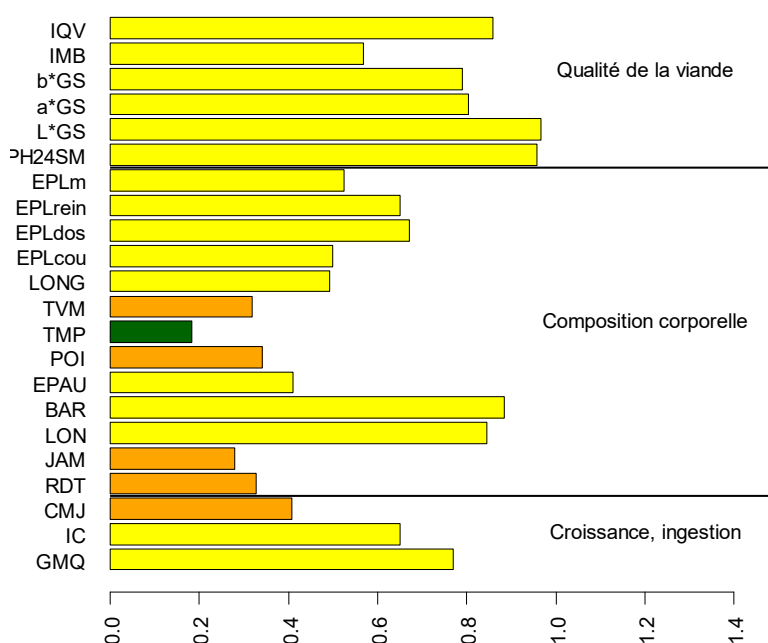


Figure 14 : Qualité de prédiction de 22 phénotypes de production à partir du métabolome plasmatique. Les erreurs moyennes de prédiction sont présentées. Ces erreurs sont moindres pour le phénotype TMP (taux de muscle estimé à partir des pièces). Trois classes de prédiction peuvent être dégagées : bonne (couleur verte), moyenne (orange) et médiocre (jaune), (Rohart et al. 2012). IQV : Indice de qualité de viande, IMB : Temps d'imbibition du muscle gluteus superficialis (10s), a*GS : indice a* de coloration rouge du muscle gluteus superficialis, b*GS indice b* de coloration jaune du muscle gluteus superficialis, L*GS : luminosité L* du muscle gluteus superficialis, pH24SM : pH 24h du muscle semimembranosus, EPLm : Epaisseur de lard moyenne (mm), EPLrein : Epaisseur de lard au niveau des reins (mm), EPLdos : Epaisseur de lard du dos (mm), EPLcou : Epaisseur de lard du cou (mm), LONG : Longueur de carcasse (mm), TVM : Teneur en viande maigre (%), TMP : Teneur en muscle estimée à partir des pièces (%), POI : Poids de la poitrine (kg), EPAU : Poids de l'épaule (kg), BAR : Poids de la bardière (kg), LON : Poids de la longe (kg), JAM : Poids du jambon (kg), RDT : Rendement de carcasse (%), CMJ : Consommation moyenne journalière (kg.j-1), IC : Indice de consommation (kg.kg-1 poids vif), GMQ : Gain moyen quotidien (kg.j-1).

L'évaluation du potentiel de prédiction des données métabolomiques a été réalisée dans le cadre de la thèse de Florian Rohart encadré par Magali San Cristobal. Ces travaux ont fait l'objet d'une collaboration avec l'organisation BIOPORC, aujourd'hui Alliance R&D, qui regroupe plusieurs organisations de sélection porcine (OSP). Les animaux étudiés sont les collatéraux des animaux utilisés par les sélectionneurs dans leurs schémas de sélection. Ces collatéraux sont élevés ensemble en croissance dans un même lieu, la station Du Rheu. Ce dispositif est un outil formidable pour comparer des lignées génétiques différentes dans un même environnement. Cette collaboration permet aussi d'évaluer le potentiel en termes d'application de l'usage de différents -omiques en lien avec la production porcine (interlocutrice Marie-Josée Mercat, IFIP/Alliance R&D).

Les cinq lignées porcines sont le Duroc, le Piétrain, le Landrace, les lignées mâle et femelle de Large White. Dans l'étude du transcriptome, il y avait aussi quelques porcs d'une lignée sino-européenne (lignée Duoshan) ; ils n'ont pas été retenus dans les études publiées mais donnent quelques indications sur ce que peut être l'expression de gènes dans les lignées chinoises souvent de caractères plus rustiques que nos lignées européennes.

Pour le transcriptome, les analyses ont été effectuées sur 147 animaux (SanCristobal et al. 2015). Les analyses différentielles (modèle linéaire et pairwise analyse ; FDR 0,1%) et discriminantes ont identifié 1230 gènes uniques. Le principal résultat pour l'analyse différentielle a été d'identifier qu'une voie de signalisation cellulaire, la voie PI3K/AKT, avait probablement été la cible de la sélection génétique recherchant à améliorer les performances des animaux. En effet, cette voie de signalisation est bien connue pour être impliquée dans la régulation de la masse musculaire ; voir les revues suivantes (Glass 2003) et (Egerman & Glass 2014). Dans notre étude, 82 gènes régulés au moins entre deux des cinq

lignées génétiques sont impliqués dans cette voie de signalisation présentée ici de façon synthétique ; 40 gènes ont été replacés dans la Figure 15. Ce résultat nous a permis de suggérer qu'en ayant sélectionné nos animaux pour une croissance musculaire accrue, la sélection génétique a pu modifier d'autres fonctions biologiques plus impliquées dans des caractères fonctionnels, en lien avec la robustesse, comme la régulation de métabolismes énergétiques (glycolyse, métabolisme lipidique). Cette voie de signalisation constitue un nœud stratégique des voies de signalisation régulant le développement et différents métabolismes. Nous y retrouvons des gènes déjà étudiés en génétique porcine : *IGF2* et *PRKAG3* (RN). Ce « nœud » moléculaire intervient dans des mécanismes aussi importants pour le développement du fœtus que pour la régulation de différents mécanismes physiologiques chez l'adulte.

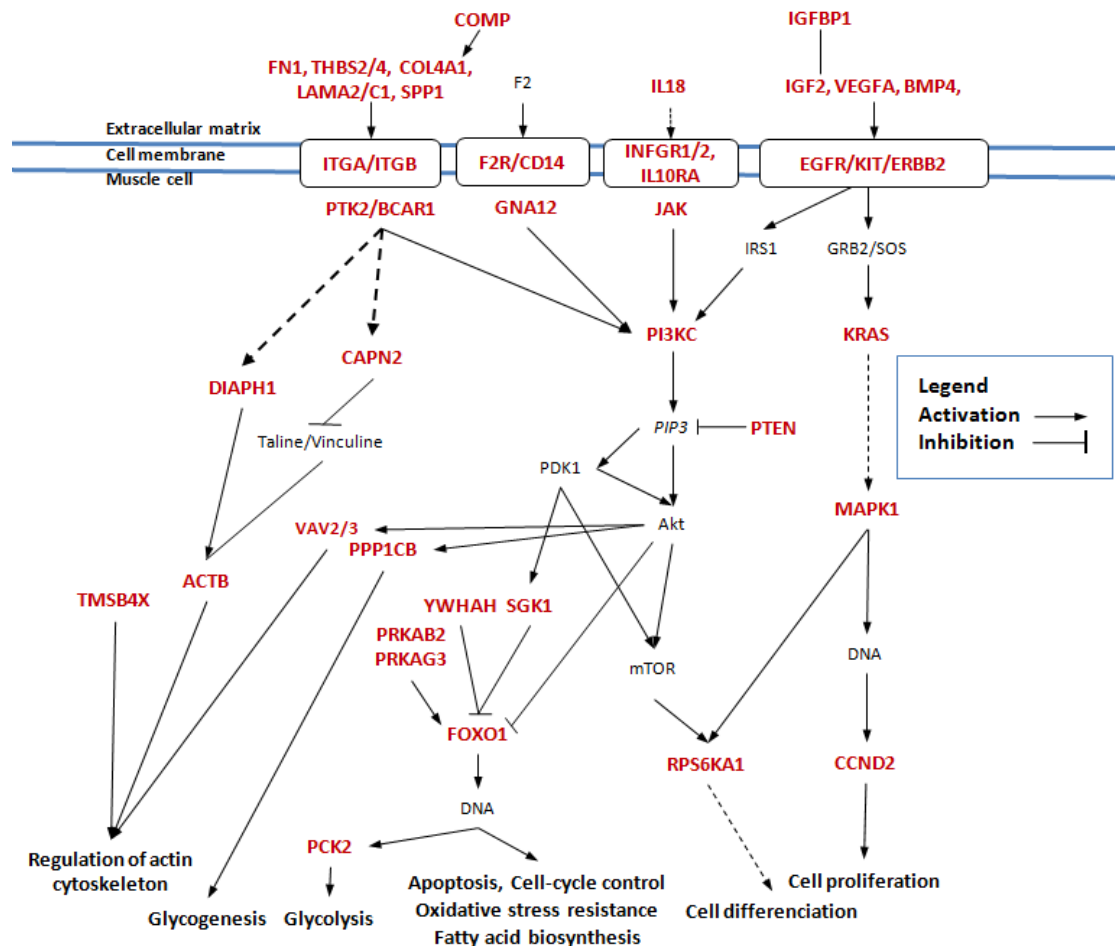


Figure 15 : Représentation synthétique de l'implication de la voie des PI3Kinases dans différents processus cellulaires. Les 40 gènes (sur les 80 identifiés) en rouge sont des gènes différenciellement exprimés entre deux des cinq lignées.

Les gènes impliqués dans ces mécanismes au centre de la régulation des phénotypes musculaires sont également expliqués dans plusieurs revues avec un focus pour l'homme mais s'appuyant aussi sur des études réalisées sur différentes espèces (Blaauw et al. 2013). Une autre revue reprend les résultats obtenus à partir des études de perte et de gain de fonction pour différents gènes sur modèle murin (Verbrugge et al. 2018). Cette revue met en particulier en avant le rôle de la voie de signalisation Igf1-Akt-mTOR.

L'autre résultat intéressant de ce travail est l'identification de 85 transcrits capables de différencier les lignées LW (mâles et femelles ensemble), Duroc, Pietrain et Landrace. Ces gènes ont été identifiés par la méthode des Forêts Aléatoires (RF, Breiman (Breiman 2001)). Les quatre lignées sont parfaitement discriminées (résultat non montré, Figure 4 du papier). Avec une autre approche discriminante, la sPLS-

DA (Le Cao et al. 2011) a permis d'identifier 147 sondes différentielles qui discriminent bien les quatre lignées. Finalement, une recherche de gènes différentiels pour une lignée comparée à toutes les autres a donné une liste de 92 sondes. La totalité de l'analyse donne 169 gènes uniques. L'intersection des trois approches permet de sélectionner 15 gènes : six gènes différencient les Duroc des autres lignées, cinq gènes pour les Piétrain et quatre gènes les LW femelles. Pour discriminer les Landrace des autres lignées, seules les méthodes des RF et la recherche de gènes spécifique de la lignée ont donné deux gènes. Si quelques gènes sont bien connus dans cette liste comme *IGF2* et *KIT*, et en dehors de la voie des PI3K, beaucoup de gènes différentiels entre les lignées sont des gènes impliqués dans des processus fondamentaux de la régulation cellulaire comme la régulation transcriptionnelle et post-transcriptionnelle. Dit autrement peu de gènes impliqués dans des voies métaboliques directement en lien avec les phénotypes terminaux ciblés par la sélection génétique sont particulièrement identifiés.

Les données du transcriptome différencient plus les Duroc des autres lignées, puis ce sont les LW qui se différencient avec les Piétrain et les Landrace en position intermédiaire (Figure 13, l'ACP pour les ARNm). Cette répartition est très proche de celle observée avec les marqueurs génétiques. Les données de traces de sélection n'ont pas été publiées. Par contre, Rubin et al. (2012) ont publié des signatures de sélection chez le porc et parmi les six gènes mis en avant par les auteurs, trois sont différentiels dans notre liste : *NR6A1* (Nuclear Receptor 6 A1, SSC1, longueur du corps), *CASP10* (Caspase 10, SSC15) et *KIT* (KIT Proto-Oncogene, Receptor Tyrosine Kinase, SSC8, pigmentation). **Ces travaux d'intégration des données d'expression avec les données génétiques (QTL, GWAS, trace de sélection) du projet Delisus restent à réaliser. L'intégration peut aussi se faire avec les données métabolomiques et entre données métabolomiques, transcriptomiques et phénotypiques. L'intégration des données fonctionnelles permettrait de comprendre comment se traduit la variabilité d'expression du génome au niveau du métabolome et des phénotypes. Ces données constitue une base d'information moléculaire et fonctionnelle qui devrait aider à élucider les mécanismes moléculaires sous-jacents les régions du génome sites de la variabilité phénotypique.**

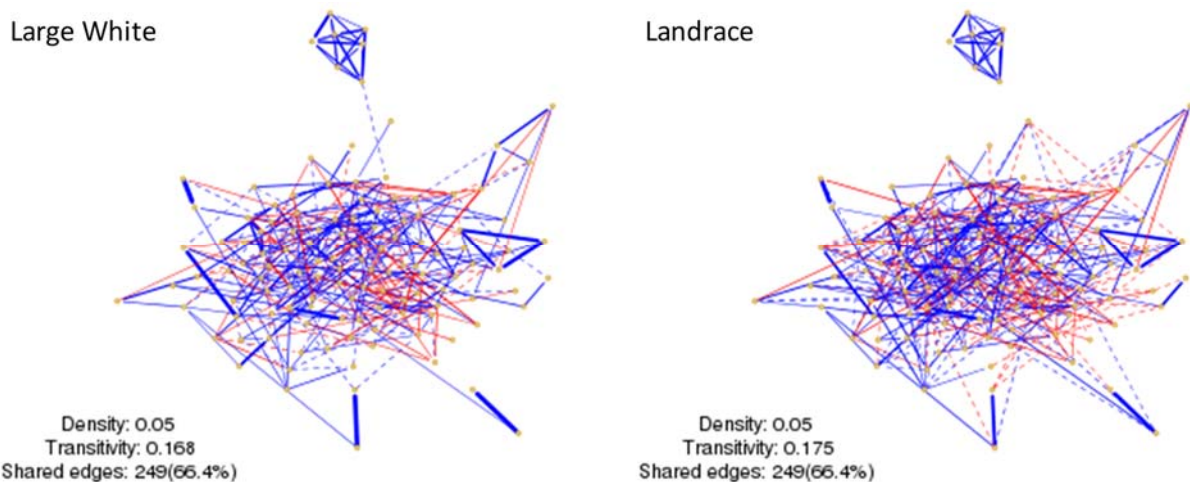


Figure 16 : Représentation de réseaux inférés spécifiques de chaque race à partir des 123 gènes les plus différentiellement exprimés chez les Large White et les Landrace.

Les liens pleins sont communs entre les deux races alors que les liens hachurés sont spécifiques de la race. La couleur bleue correspond à une corrélation partielle positive et la couleur rouge correspond à une corrélation partielle négative.

Une autre façon d'utiliser les données d'expression est d'inférer un réseau spécifique à chaque race et ensuite de comparer les réseaux entre eux. Un premier essai effectué par Magali San Cristobal avec Nathalie Vialaneix donne les réseaux de la Figure 16 où 66 % des liens sont conservés entre les deux lignées. **Ensuite, les études à faire sont celles réalisées dans les autres projets ; statistiques spatiales, identification de nœuds avec un fort degré ou une forte centralité, recherche de communautés, impact fonctionnel. Ainsi, parmi le gènes testés, peuvent être identifiés ceux qui ont un rôle plus ou moins important dans la structure du graphe, l'impact sur la répartition des gènes en communautés, l'impact sur des fonctions biologiques enrichies.**

3. Etude d'un caractère complexe par une approche intégrative : la tendreté de la viande

Mon premier projet de recherche à mon arrivée à l'INRA était un projet sur la qualité de la viande et en particulier sur la tendreté de la longe de porc. Ce projet était très ambitieux par rapport à la disponibilité des outils nécessaires à sa réalisation. Il s'agissait d'un projet de génétique génomique où la transcriptomique s'inscrivait dans un contexte génétique. Ce projet a été une initiation à la génétique en abordant des aspects qui émergeaient en 2001. Il s'agissait donc d'associer de la génétique quantitative à l'analyse du transcriptome (Jansen & Nap 2001 ; Brem *et al.* 2002; Kadarmideen *et al.* 2006). Sachant bien que mon rôle était d'analyser le transcriptome, pas de faire de la génétique quantitative.

Ce projet, financé par le ministère de la recherche (AG-PSG D116), était coordonné par Pierre Cherel (France Hybride racheté depuis par Hendrix Genetics) et impliquait plusieurs laboratoires INRA avec Pascale Le Roy (consolidation statistique des données et approches eQTL), avec G. Monin et E. Laville (protéomique), avec D. Milan (développement de marqueurs génétiques pour les typages), F. Hatey et moi-même (étude du transcriptome), avec M. Damon (étude du transcriptome). La partie transcriptome de ce projet était donc partagée entre deux laboratoires (LGC aujourd'hui GenPhySE et SENAH aujourd'hui PEGASE), Marie Damon étudiant la déstructuration du jambon et moi la tendreté de la longe. Ce travail a bénéficié du soutien d'une stagiaire postdoctorale, Valérie Lobjois.

a. Recherche de gènes régulés en fonction d'un phénotype quantitatif.

Les qualités sensorielles et nutritionnelles de la viande dépendent de différents facteurs physiologiques (nombre, type et taille des fibres musculaires), environnementaux (conditions d'élevage, nutrition, transport, abattage, transformation) et génétiques (Morzel *et al.* 2004; Picard *et al.* 2015; Listrat *et al.* 2016). Dans ce contexte, le but du projet D116 « Recherche de prédicteurs précoces et de QTL associés aux défauts de qualité de la viande de porc à travers l'analyse du transcriptome » était de mieux caractériser la variabilité génétique en relation avec les critères de qualité de la viande à travers l'analyse du transcriptome du muscle strié squelettique. Les critères de qualité ciblés étaient la tendreté de la longe (*Longissimus dorsi*, LD) et la déstructuration du jambon (*Semi membranous*, SM).

La qualité de la viande de porc dépend des caractéristiques physico-chimiques du muscle. Parmi les qualités sensorielles, la tendreté apparaît comme un critère important du point de vue des consommateurs (Maltin *et al.* 2003). Les études réalisées chez le porc montrent que différentes caractéristiques histologiques (type des fibres, longueur des sarcomères) et biochimiques (contenu et type de collagène, protéolyse) peuvent contribuer à la variabilité de la tendreté (Wheeler *et al.* 2000; Picard *et al.* 2010). Cependant, les mécanismes moléculaires impliqués dans cette variabilité restaient à identifier.

Les échantillons utilisés dans le projet D116 provenaient d'une collection de plus de 2000 échantillons de muscles (LD et SM) prélevés juste après abattage, provenant de 1100 animaux répartis en 19 familles QTL. A cette collection est associée une base de données contenant les informations sur la qualité de la viande fraîche et transformée correspondant à ces échantillons (pH, résistance au cisaillement, colorimétrie, potentiel glycolytique, lipides intramusculaires...). La Figure 17 présente le dispositif et l'ensemble des phénotypes mesurés dans ce projet.

Production des animaux expérimentaux F2 :
rassembler des facteurs de variation d'origine génétique

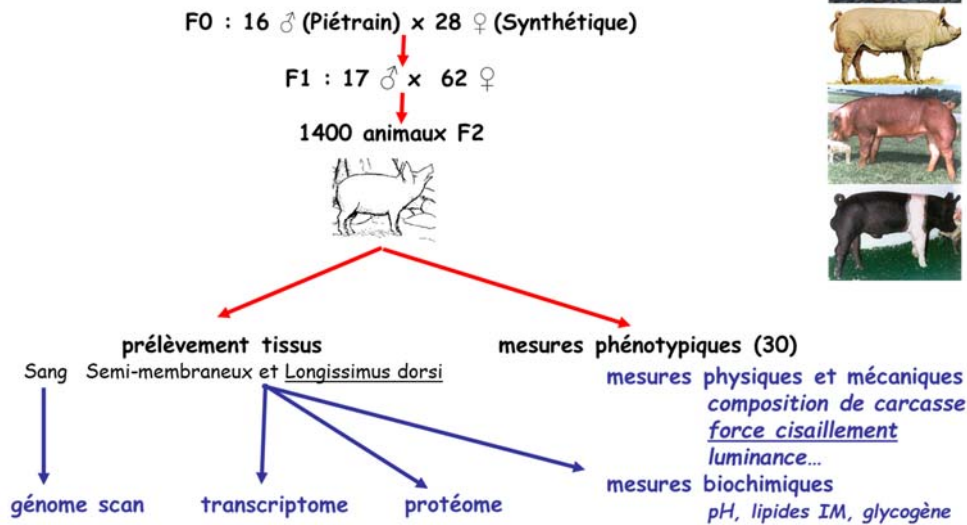


Figure 17 : Présentation du dispositif expérimental pour l'analyse du transcriptome du muscle dans un contexte génétique.

Les porcs utilisés étaient des animaux F2 produits à partir de deux lignées mâles commercialement utilisées pour la production de verrats terminaux (lignée FH016, de type Piétrain, et lignée FH019, lignée synthétique issue de fondateurs Large White, Duroc et Hampshire, France Hybrides). Le Longissimus dorsi est le muscle étudié pour le caractère ciblé qui est la tendreté et le phénotype mesuré est la force de cisaillement.

Les informations de génotypage de chacun des animaux (120 marqueurs de type microsatellite couvrant l'ensemble du génome) sont disponibles pour la primo-localisation de QTL Cherel (Cherel et al. 2011). L'objectif était d'utiliser les données d'expression pour une caractérisation plus fine des phénotypes afin de rechercher des QTL d'intérêt et aussi d'orienter les produits vers une utilisation plus adaptée (viande fraîche ou produit transformé). Dans le cadre de ce projet de génomique comprenant des analyses génétiques (Cherel *et al.* 2011), du protéome (Sayd et al. 2006; Laville et al. 2007; Laville et al. 2009) et du transcriptome (Lobjois et al. 2008; Liaubet et al. 2011; Cherel et al. 2012) ont été réalisées.

L'analyse du transcriptome de la longe s'est révélée bien plus complexe qu'initialement prévue pour le caractère tendreté mesuré par la force de cisaillement. Il était d'abord envisagé de ne faire que la comparaison des échantillons considérés extrêmes pour le phénotype mesuré. Les échantillons avaient été choisis pour être extrêmes dans l'une des familles QTL donc 30 échantillons pour 15 familles. Suivant ce modèle, seulement 33 sondes se sont avérées différentielles pour un FDR de 20% ou 13 sondes pour un FDR de 5%.

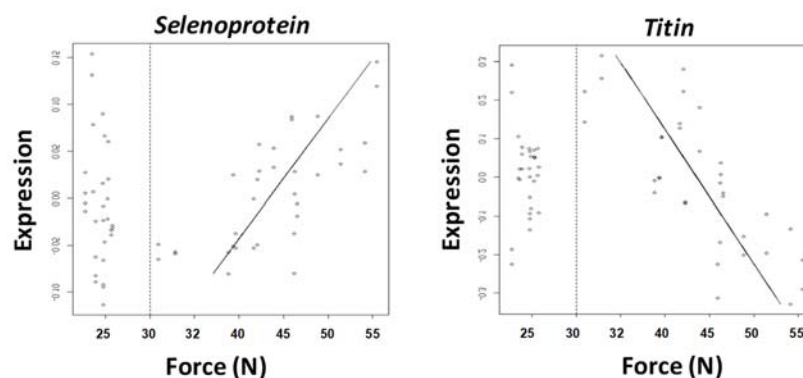


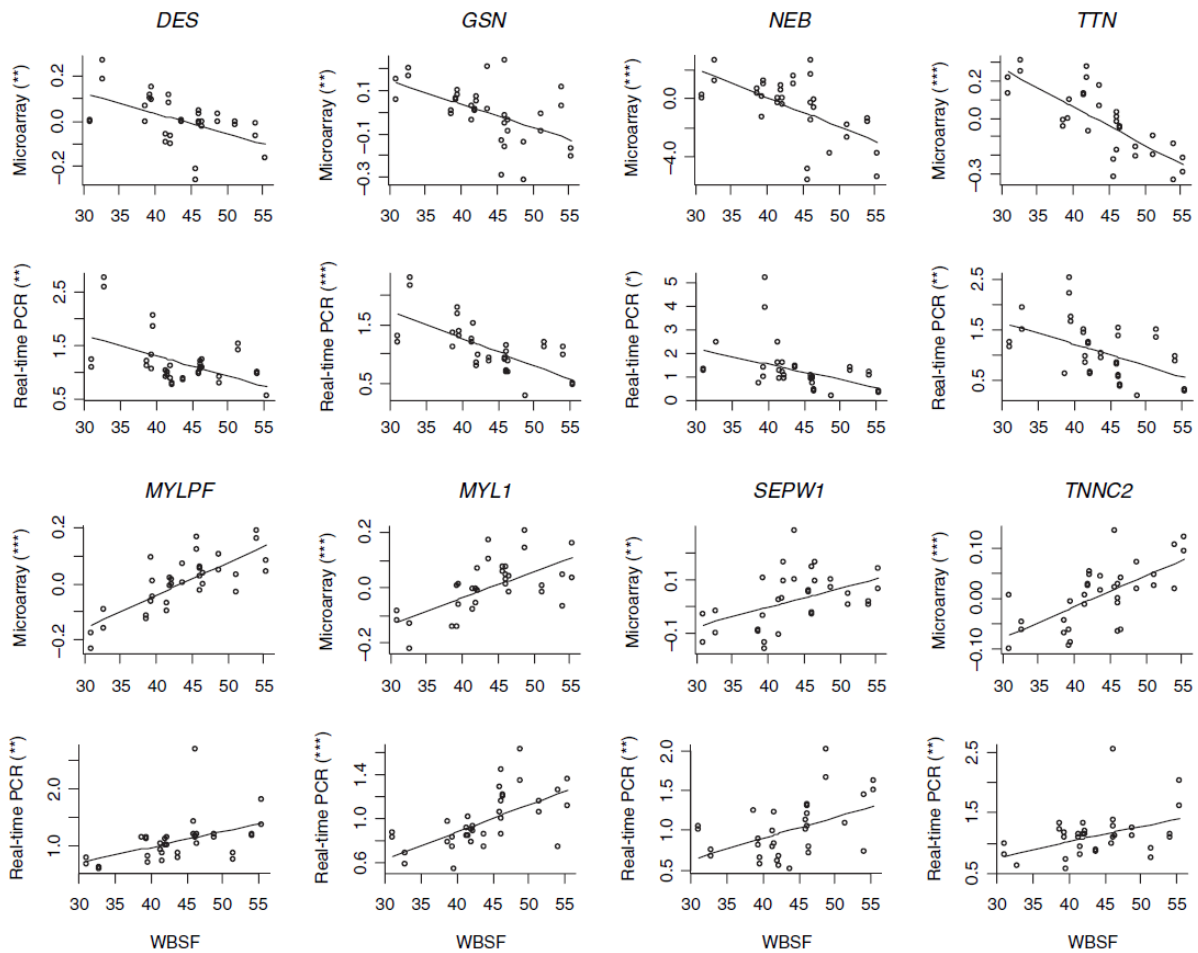
Figure 18 : Expression de deux gènes (SELW et la TTN) en fonction des valeurs de forces de cisaillement mesurées sur viande cuite.

On observe une relation linéaire pour les variations d'expression des gènes que pour les valeurs de force de

cisaillement supérieures à 30 Newtons.

L'exploration des données a montré que les variabilités d'expression n'étaient pas linéaires (Figure 18). Les données n'étaient alors exploitables qu'en étudiant séparément les échantillons ayant une force de cisaillement > 30 N des autres. Nous n'avons aucune explication à ce phénomène, mais Wheeler et al. (2000) avaient observé un phénomène assez similaire : une note maximale de tendreté était obtenue dans leur expérience jusqu'à une certaine longueur du sarcomère de 2 μm indépendamment de la teneur en collagène et de l'activité protéolytique.

Dans ces conditions et après vérification des hypothèses inhérentes au modèle de régression linéaire, nous avons identifié 118 clones significatifs ($P < 0,001$; taux de faux positifs $< 0,05$) pour lesquels le signal d'hybridation est corrélé avec la force de cisaillement. Ces clones correspondent à 65 gènes après annotation manuelle.



**Figure 19 : Régression linéaire entre les valeurs d'expression et la force de cisaillement de huit gènes obtenues par deux méthodes différentes (microarray vs. qPCR).
Seuil de significativité (P-value: * <0.1 , ** <0.01 , *** <0.001).**

Les 118 clones se partagent en 2 groupes corrélés positivement ou négativement avec la force de cisaillement. Nous retrouvons dans cette analyse des gènes du cytosquelette et du métabolisme mais aussi des facteurs de transcription ou des médiateurs cellulaires. L'analyse des variations d'expression de 8 gènes pertinents (Figure 19) par RT-PCR quantitative donne les mêmes résultats.

Le logiciel IPA (<http://www.ingenuity.com>) a été utilisé pour faciliter l'annotation fonctionnelle des gènes ; ce logiciel de fouille des données bibliographiques identifie tout type de lien qui peut exister entre les gènes différentiels et d'autres gènes, un test statistique estime si la représentation des gènes

différentiels est significative (enrichissement) dans le réseau proposé. Ainsi, 35 gènes ont été regroupés dans 3 réseaux de gènes (Lobjois *et al.* 2008). La Figure 20 présente un nouveau réseau réalisé deux ans plus tard. Les données bibliographiques utilisées par IPA pour construire les réseaux ont évolué. Le logiciel étant interactif, tous les liens peuvent être vérifiés en retournant aux publications. Ce nouveau réseau inclut 23 gènes différentiellement impliqués dans le développement et la morphologie du tissu musculaire. Les réseaux de gènes proposent aussi des liens avec des gènes dont les modalités d'expression ne sont pas connues dans notre système ; ici, parce que les gènes sont absents de la puce porcine utilisée.

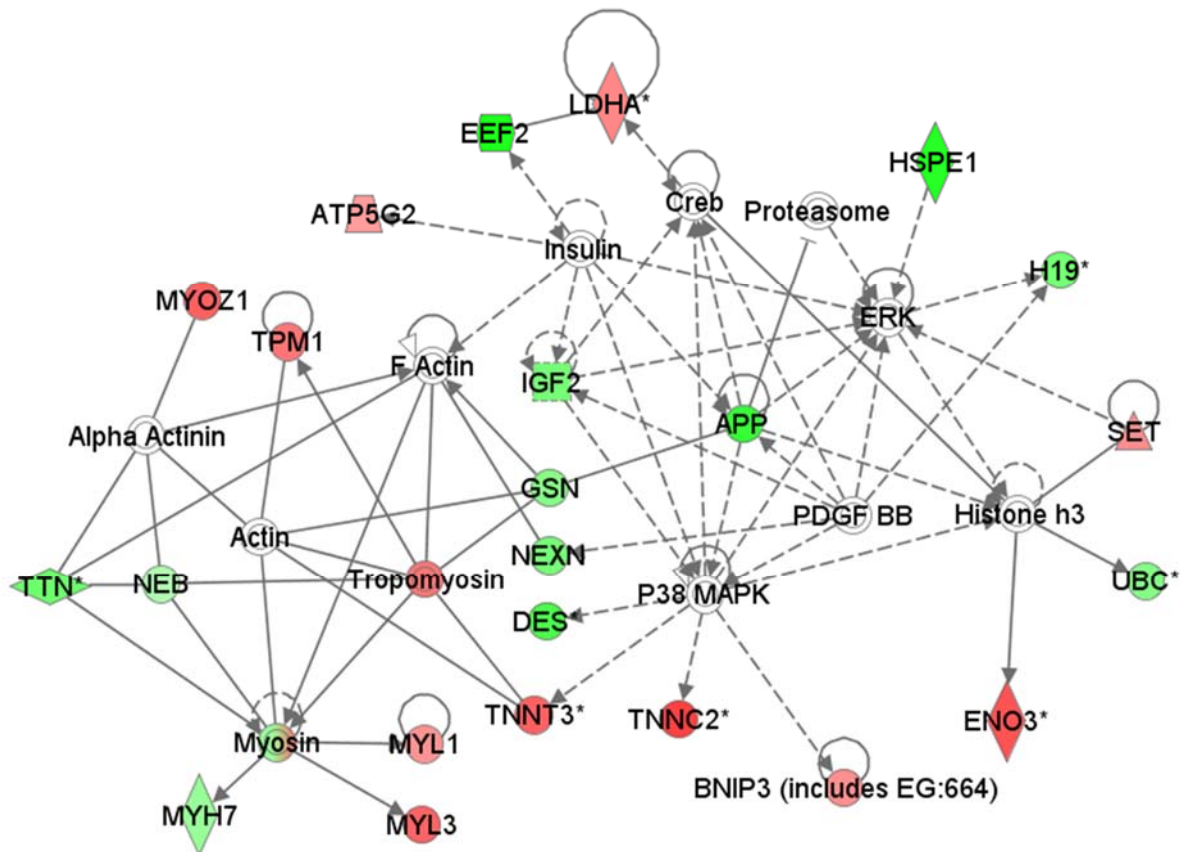


Figure 20 : Réseau bibliographique obtenu avec le logiciel Ingenuity incluant 23 gènes différentiellement impliqués dans le développement et la morphologie du tissu musculaire. Les gènes en vert ont une expression décroissante et en rouge croissante en fonction de la force de cisaillement.

Les résultats d'enrichissements significatifs peuvent ensuite être étudiés pour interprétation. Cette interprétation peut évoluer à la marge, avec l'amélioration de l'annotation et aussi l'amélioration de l'information acquise sur les gènes, les ontologies. Concrètement, ce qu'on observe est plutôt une amélioration de l'enrichissement des ontologies et des scores associés aux réseaux déjà obtenus.

b. Intégration des données issues de l'analyse du transcriptome et du protéome.

D'autre part, la comparaison et l'intégration des données issues des analyses du transcriptome et du protéome permettent d'affiner la compréhension des mécanismes moléculaires de la tendreté chez le porc. Les deux types de données ont été étudiés séparément et je n'ai pas du tout été impliquée dans l'obtention et l'analyse des données protéomiques (Laville *et al.* 2007). L'analyse du protéome s'est faite par électrophorèse 2D après une étape d'extraction des protéines cytoplasmiques et en excluant les protéines du cytosquelette. L'analyse statistique a été faite en classe (les deux groupes de faibles et fortes valeurs de force de cisaillement, deux échantillons par famille). Laville *et al.* (2007) ont identifié 14 spots différentiels pour une ($P < 0.05$). Entre les deux jeux de données, un seul gène différentiel est également retrouvé différentiel au niveau protéique. Il s'agit de l'Adénylate Kinase (gène *AK1*). L'Adénylate Kinase catalyse le transfert réversible du groupe phosphate terminal entre l'ATP et l'AMP,

cette enzyme joue un rôle important dans l'homéostasie cellulaire. L'intégration des données protéomiques et transcriptomiques a été initiée mais pas finalisée et donc pas publiée. Les essais réalisés n'ont pas relevé d'une approche intégrative mathématique puisque les échantillons retenus pour l'analyse finale ne sont pas complètement les mêmes et le choix technique d'exclure les protéines du cytosquelette permet de cibler qu'une partie du protéome cytoplasmique. Le choix a donc été d'utiliser le logiciel IPA pour combiner les deux types de données et observer leur complémentarité.

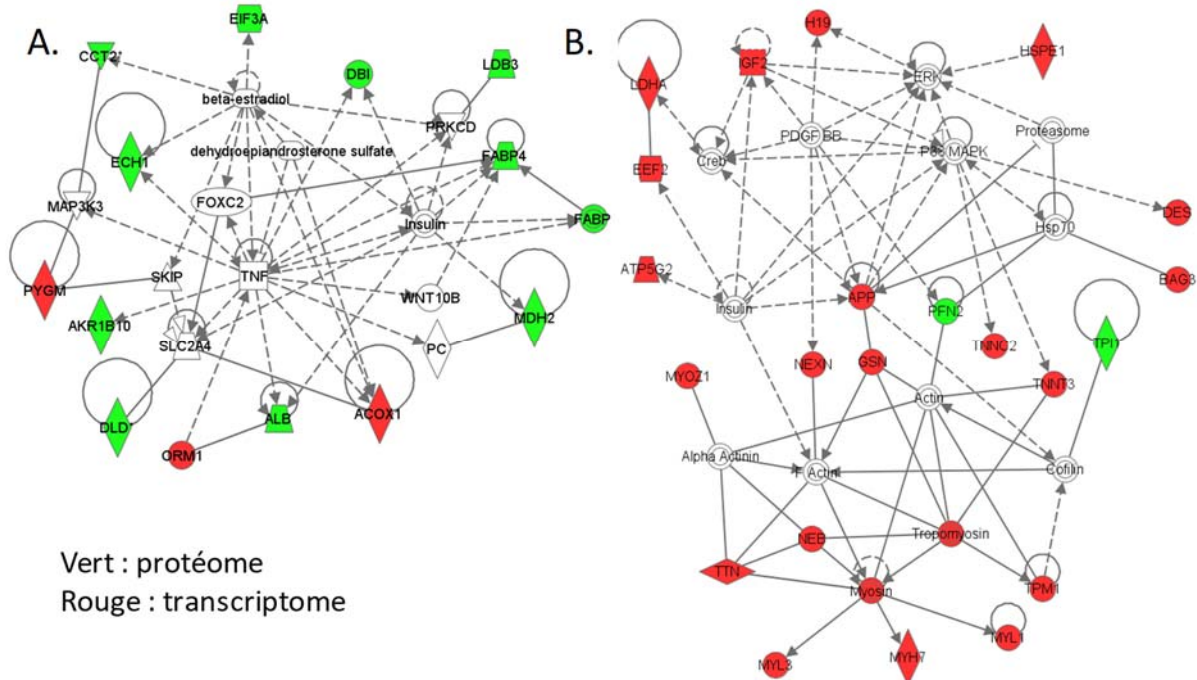


Figure 21 : Analyse des données du protéome et du transcriptome du muscle Longissimus par une approche de réseaux de co-occurrence (logiciel IPA).

Deux réseaux significatifs ont été obtenus. A. Le premier réseau inclut surtout des protéines. B. Le deuxième réseau implique surtout des gènes différentiels pour des valeurs de forces de cisaillement supérieures à 30 Newtons.

Les deux jeux de données concaténés en un seul ont été soumis à IPA. La Figure 21 montre que le logiciel propose deux réseaux mais finalement avec une certaine séparation fonctionnelle des résultats du protéome et du transcriptome. Le protéome (Figure 21A) semble plus expliquer le métabolisme lipidique impliqué dans les valeurs basses de force de cisaillement (Laville *et al.* 2007) alors que la morphologie musculaire s'explique ici plus avec les données du transcriptome (Figure 21B) et pour des forces de cisaillement élevées. La combinaison des deux approches complète la description des mécanismes biologiques sous-jacents le caractère. Toutefois, même dans ce cas, les deux outils – omiques utilisés sont encore loin d'explorer l'ensemble du transcriptome et du protéome.

c. Organisation génomique des gènes régulés.

Nous avons localisé les gènes précédemment identifiés pour la tendreté sur le génome du porc. Le génome porcin n'était alors pas encore séquencé, la position sur le génome du porc était déterminée avec les outils de génomique comparative homme/porc pour prédire la localisation génomique porcine en utilisant les différentes cartes disponibles. Nous avons identifié 20 gènes de localisation prédite intéressante pour une possible co-localisation avec des QTL et nous les avons localisés sur la carte RH en utilisant le panel d'hybrides irradiés (Lobjois *et al.* 2008; Jouffe *et al.* 2009)).

Nous observons que des QTL pour la tendreté et la force de cisaillement (Figure 22), données PigQTLdb (Hu *et al.* 2005) et de travaux originaux (Malek *et al.* 2001; Hu *et al.* 2005; Rohrer *et al.* 2006)) co-localisent avec des gènes différentiellement exprimés. Ces gènes sont donc des candidats positionnels. Cette démarche fait échos à celle de mes collègues généticiens moléculaires qui, lorsqu'ils ont des régions QTL identifiées, recherchent quels gènes co-localisés pourraient avoir une fonction liée au

caractère étudié. Le premier test qu'effectuent les généticiens moléculaires est généralement de tester si ces gènes sont différentiellement exprimés en fonction des allèles et du phénotype. En effectuant la démarche inverse, il est possible d'observer que les gènes en lien avec le caractère étudié peuvent être des gènes sans *a priori* fonctionnel lié au caractère voire même des gènes peu ou pas caractérisés. Ainsi, 12 gènes (Figure 22, (Lobjois *et al.* 2008)) ont été localisés dans 4 régions chromosomiques d'intérêt sur les chromosomes 2 (Malek *et al.* 2001; Rohrer *et al.* 2006), 6 (Szyda *et al.* 2003) et 13 (de Koning *et al.* 2001). **L'utilisation du transcriptome permet de proposer des gènes qui n'auraient pas forcément retenus l'attention sur la seule base de leur fonction connue.**

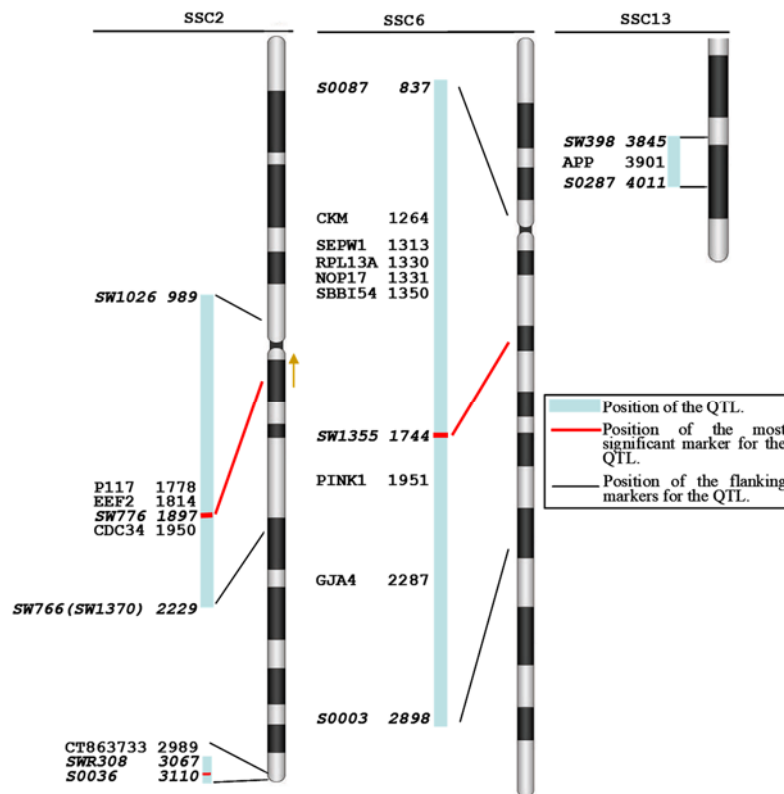


Figure 22 : Co-localisation entre des gènes différentiellement exprimés et des QTL pour le tendreté de la viande. La recherche de la co-localisation des gènes différentiels avec des QTL publiés dans le cadre d'autres projets de recherche permet de proposer des gènes comme candidats positionnels.

La démarche utilisée pour la recherche de co-localisation a aussi été utilisée dans un autre travail collaboratif au niveau européen dans le cadre du programme EADGENE (Jouffe *et al.* 2009) où l'objectif était de confronter nos démarches d'analyse des données d'expression dans un contexte génétique. Ce qui était alors intéressant, c'était que certains avaient déjà accès aux premières données de séquençage, d'assemblage et d'annotation (Figure 23). L'intégration des données QTL et des données d'expression pour des espèces non modèles comme le porc est très difficile en l'absence d'une séquence complète du génome. Utiliser la dernière version disponible du génome porcine offre la voie la plus directe pour localiser les gènes différentiellement exprimés ainsi que les marqueurs génétiques flanquant les QTLs. Pour ce travail, il s'agissait des versions 6 et 7 du génome porcine, la version actuellement disponible est la Sscrofa11.1. Pour les versions 6 et 7, il ne s'agissait que d'une séquence de pré-Ensembl. Ce qui signifie que l'assemblage était encore incomplet et susceptible de contenir un grand nombre d'erreurs. Par conséquent, l'emplacement des gènes et des marqueurs utilisant ce brouillon n'était en aucun cas une donnée sûre. Même pour le nombre limité de gènes utilisé dans ce travail, les différentes façons de les localiser pouvaient engendrer des divergences. Un bon exemple était le cas du gène *RNF2* (Ring Finger Protein 2) qui s'alignait directement sur le chromosome 1 où justement était localisé un QTL d'intérêt. J'avais alors démontré que sur le chromosome 1, il y avait un pseudogène de *RNF2* et que le gène codant était, *a priori* par synténie, sur le chromosome 9. En mars 2019, *RNF2* est bien localisé sur le chromosome 9 et pas le 1. J'avais donc bien identifié l'emplacement réel du gène

contrairement à mes co-auteurs. Cette étude était très informative sur la problématique posée, identifier des gènes candidats à partir de données d'expression. La question de la qualité de l'alignement des séquences exprimées (séquences d'ADNc ou de RNAseq) sur la séquence génomique me semblent être toujours d'actualité. Je retrouve cette problématique quand il faut réannoter les sondes des différents microarrays utilisés jusqu'à présent. Je n'ai pas encore identifié aujourd'hui de procédure complètement satisfaisante pour annoter des séquences exprimées hors annotation manuelle. C'est une question que j'ai partagée avec des collègues bioinformaticiens et qui a donné lieu à deux co-encadrements de stagiaires niveau licence en 2010 (avec Thomas Faraut, GenPhySE) et en 2014 (avec Christophe Klopp, MIA-T/Sigenae).

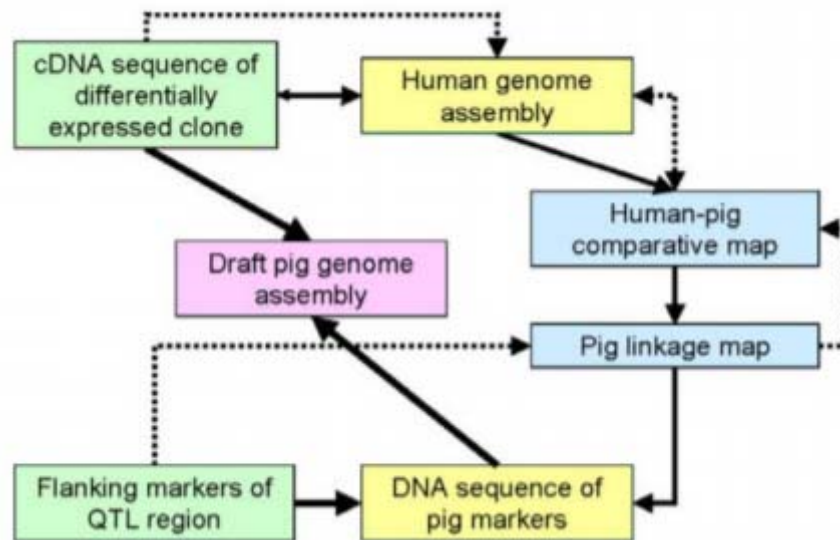


Figure 23: Schéma résumant deux stratégies d'intégration de données d'expression et génétiques dans le cadre d'un réseau européen d'analyse des données transcriptome EADGENE (Jouffe et al. 2009).

Au final, l'objectif est d'identifier des gènes potentiellement intéressants pour une recherche de polymorphismes qui pourraient être au moins associés au phénotype d'intérêt, au mieux identifiés comme étant causal pour les variations phénotypiques. Dans le cadre du projet D116, un des gènes candidats expressionnel et positionnel de la Figure 22 s'est aussi avéré être le plus prédictif du phénotype de tendreté et pour lequel nous avons identifié un SNP... Ce résultat sera présenté et discuté dans la partie 3.e.

d. Régulation génétique de l'expression (eQTL).

Les phénotypes musculaires comme la plupart des phénotypes quantitatifs sont des traits complexes résultant de l'interaction entre les variations génétiques et environnementales. En génétique animale, de nombreux loci pour des caractères quantitatifs (QTL) ont été décrits en tant que composante de la variabilité génétique de la physiologie, de la résistance aux maladies et des caractères de reproduction (Rothschild 2003). Mais le clonage positionnel des polymorphismes sous-jacents ces QTL est un long processus, de même que la compréhension des cascades moléculaires conduisant aux phénotypes observés. L'identification des QTL d'expression (eQTL) permet de caractériser les premiers effets de la variabilité génétique et ainsi permettre de comprendre les processus moléculaires affectés. Les gènes dont les transcrits sont affectés par ces eQTL ne sont pas forcément porteurs de la mutation causale à la variation d'expression observée mais la distribution des localisations des eQTL permet de définir des groupes fonctionnels de gènes potentiellement co-régulés génétiquement (Rothschild 2004; Liaubet et al. 2011). Les variations de l'expression des gènes seraient responsables d'une grande partie de la variation phénotypique observée chez les populations naturelles. Des changements dans la régulation des gènes se sont avérés sous-jacents à des phénotypes adaptatifs chez différentes espèces (Oleksiak et al. 2002; Gilad et al. 2006). Pour toutes ces raisons, les études de cartographie eQTL constituent un

nouvel outil puissant pour identifier les variants génétiques régulant l'expression des gènes (Jansen & Nap 2001; Brem *et al.* 2002; Schadt *et al.* 2003; Yvert *et al.* 2003). L'analyse du transcriptome mesure le niveau d'expression de nombreux gènes qui sont les phénotypes dans l'analyse eQTL et la ségrégation des marqueurs génétiques au sein des familles permet la cartographie des loci affectant ces phénotypes à des régions génomiques spécifiques. Les analyses globales eQTL ont permis la détection de la variation génétique *cis* contrôlant des gènes individuels et de nombreux eQTL *trans* regroupés qui régulent donc un ensemble de gènes (Gilad *et al.* 2008).

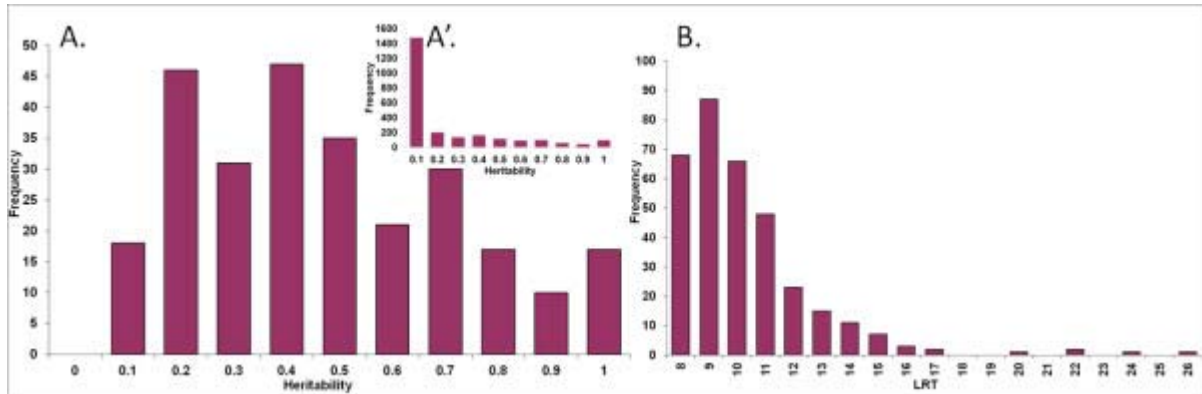


Figure 24 : Distribution des estimations de l'héritabilité (A) pour les 272 gènes avec eQTL et (A') pour les 2464 gènes exprimés dans le muscle et (B) l'histogramme des valeurs de LRT des 272 gènes.

Donc, pour aller plus loin dans l'intégration des données d'expression et des données génétiques, une famille F2 de 57 porcs du projet D116 a été choisie pour rechercher des eQTL. Dans cette étude, l'héritabilité des gènes exprimés a été estimée (Figure 24). L'héritabilité des niveaux d'expression a été estimée pour sélectionner les transcrits pour lesquels le niveau d'expression semble régulé génétiquement. La Figure 24A' montre l'héritabilité calculée pour tous les transcrits. Une héritabilité d'au moins 5% a été estimée pour 1057 transcrits, qui ont ensuite été analysés pour la détection des eQTL. La grande proportion des transcrits pour lesquels la variance estimée n'était pas significative (57%) reflète des transcrits présentant une variance faible ou nulle, ou pour lesquels la puissance du dispositif est trop limitée pour l'estimer. L'héritabilité moyenne pour les 1057 transcrits dont l'expression est jugée héréditaire ($h^2 > 0,05$) était de 0,18 (Figure 24A), cette héritabilité moyenne est de 0,45 pour les 272 transcrits avec eQTL.

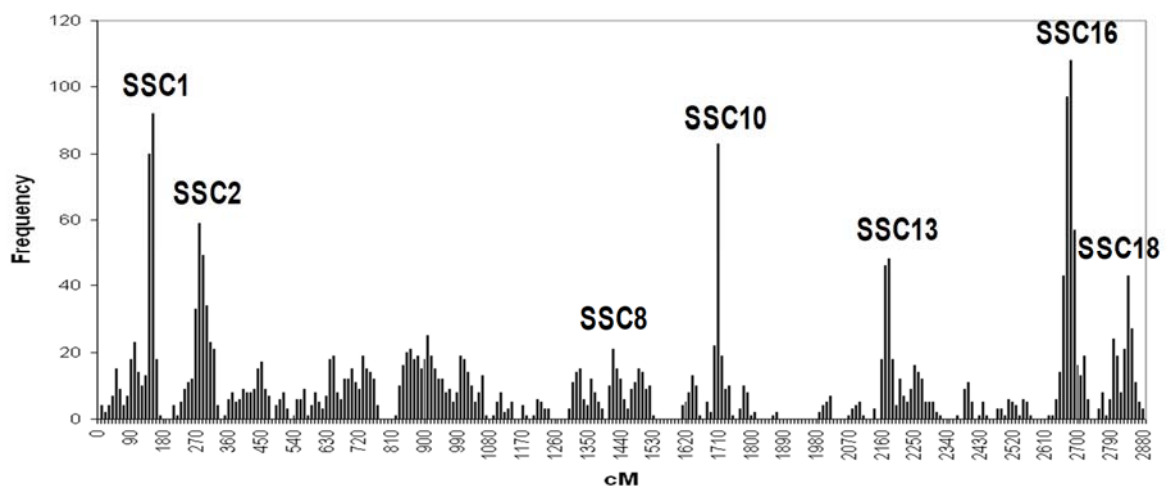


Figure 24 : 335 régions génomiques (eQTL) identifiées pour 272 transcrits. La moitié des eQTL sont regroupés en clusters de trans-eQTL pour lesquels nous pouvons parfois identifier un cis-eQTL putatif. Ce cis-eQTL devient un gène candidat positionnel pour expliquer les effets observés au QTL et supposer un rôle régulateur (master cis-eQTL) sur l'expression des autres gènes (trans-eQTL) du cluster.

L'héritabilité de l'expression des gènes a été très peu explorée dans les études eQTL chez le porc, à l'exception de notre étude où l'héritabilité a été utilisée comme critère de filtrage pour hiérarchiser les gènes et dans une étude très récente (Velez-Irizarry et al. 2019) où les auteurs précisent que les héritabilités observées pour les gènes avec un eQTL distant étaient significativement inférieures à celles des gènes localement régulés.

Les génotypes des 57 individus ont été obtenus avec 170 marqueurs microsatellites et les profils d'expression ont été obtenus à partir des 57 longes correspondantes. Au total, 272 gènes ont été identifiés comme étant génétiquement régulés par 335 eQTL (Liaubet *et al.* 2011). La moitié de ces eQTL sont organisés en clusters sur les chromosomes 1, 2, 10, 13, 16 et 18 suggérant qu'un ou plusieurs gènes localisés dans la région génomique du cluster peuvent être responsables de la variabilité d'expression des gènes ayant des eQTL co-localisés (Figure 25).

La plupart des eQTL identifiés sont des *trans*-eQTL (les gènes sont localisés loin de leur région régulatrice), les autres (au nombre de 18) étant des *cis*-eQTL (le gène co-localise avec sa région eQTL). Attention, ici les eQTL sont déclarés *cis* selon des critères aussi précis que possible au moment où l'analyse a été réalisée mais cette caractérisation reste très hypothétique au niveau de la résolution retenue.

A titre d'exemple, le cluster de 31 eQTL localisé sur le chromosome 1 (Figure 26) contient un possible *cis*-eQTL régulant l'expression du gène *COQ4* (Coenzyme Q4) mais qui pourrait aussi réguler indirectement les 30 autres gènes du cluster dont les expressions corréleront presque toutes avec celle de *COQ4*. Le gène *COQ4* code une des protéines du complexe des Coenzymes Q impliqué dans la chaîne respiratoire mitochondriale (Turunen et al. 2004; Casarin et al. 2008). Chez le porc, plusieurs QTL régulant la croissance et l'adiposité ont d'ailleurs la même localisation que ce cluster de eQTL (Rohrer & Keele 1998; de Koning et al. 1999; Bidanel et al. 2001; Sanchez et al. 2006) suggérant que le cluster de eQTL situé sur le chromosome 1 porcin pourrait réguler la physiologie musculaire en modulant le fonctionnement de la chaîne respiratoire mitochondriale et les processus de mort cellulaire. La région génomique sur le chromosome 1 est donc une région à effet pléiotropique, une région qui affecte plusieurs phénotypes d'intérêt en modifiant l'expression de plusieurs gènes. Le polymorphisme causal reste toutefois à identifier.

Parmi les 272 gènes dont l'expression est régulée par au moins un eQTL, 5 gènes (*ACOX1*, *FKBP8*, *LDHA*, *NEB*, *SEPWI*) sont aussi des gènes dont l'expression est corrélée avec la force de cisaillement sur viande cuite au sein du même dispositif expérimental (Lobjois *et al.* 2008). Il n'est pas identifié de *cis*-eQTL pour ces gènes.

Dans la Figure 25, la corrélation génétique a été utilisée pour conforter l'hypothèse d'avoir identifié un cluster de eQTL donc avec un effet pléiotrope supposé de cette région QTL pour co-réguler l'expression d'une trentaine de gènes. Bien entendu, ici les modalités moléculaires, fonctionnelles de la co-régulation ne sont pas connues.

On notera que la recherche d'eQTL a été réalisée pour tous les gènes pour lesquels nous avons détecté une héritabilité de l'expression. Il est assez commun par ailleurs d'effectuer d'autres types de présélection, par exemples des gènes corrélés ou différenciellement exprimés pour un phénotype. Donc nos 272 gènes n'ont pas été sélectionnés pour partager une implication particulière dans un de nos phénotypes d'intérêt. Il était donc important pour nous d'essayer de faire la démarche inverse et d'explorer nos données d'expression pour lesquels nous savons juste qu'ils sont régulés par au moins un eQTL. C'est dans ce cadre que nous avons recherché à inférer un réseau de co-expression (Figure 26, (Villa-Vialaneix *et al.* 2013)).

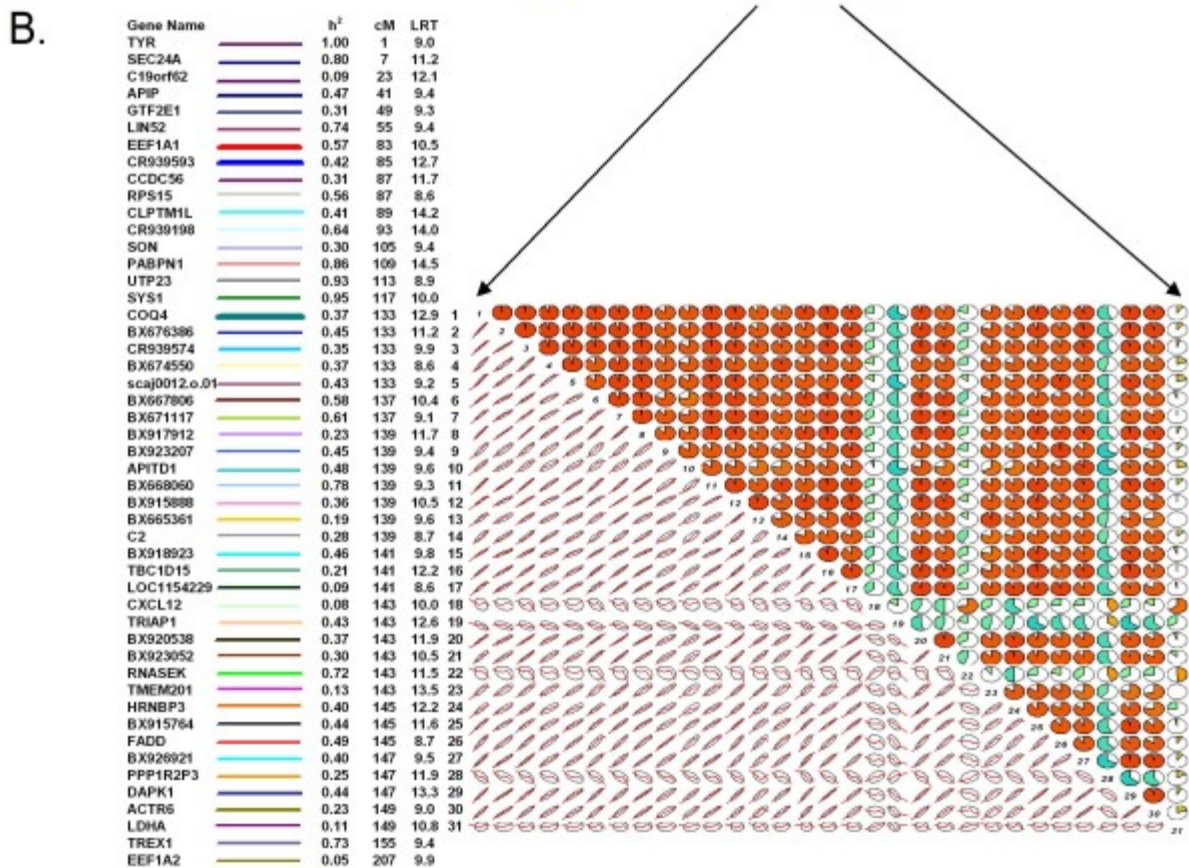
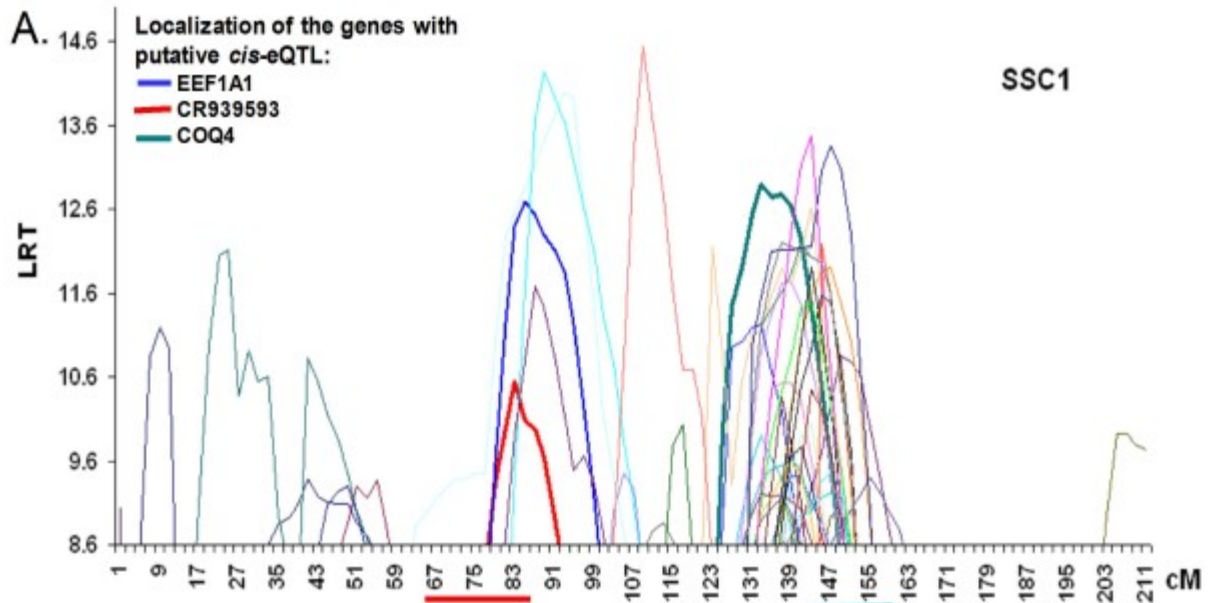


Figure 25 : Cartographie eQTL sur le chromosome 1 (A) et corrélogramme entre les gènes localisant entre 133-149 cM sur le chromosome 1 (B).

A. Le seuil chromosomique de 1% (basé sur des simulations) correspond à la valeur de LRT la plus basse de 8,6 sur le graphique. Les 49 eQTL sur SSC1 sont localisés. Les trans-eQTL sont représentés par des lignes étroites, tandis que les trois cis-eQTL possibles sont représentés par des lignes épaisses (COQ4, EEF1A1 et CR939593). B. Le corrélogramme comprend les 31 gènes avec eQTL dans le groupe 133-149 cM avec les lignes colorées utilisées pour la cartographie. Le corrélogramme comprend un cercle et une ellipse. Chaque cercle est ombré en rouge ou en vert en fonction du signe (+ et - respectivement) de la corrélation et de l'intensité de la couleur graduée de 0 à 100% proportionnellement à l'ampleur de la corrélation. L'ellipse est une matrice de diagramme de dispersion schématique et chaque panneau décrit les modèles de relations entre les variables avec une ellipse de confiance et une courbe lissée.

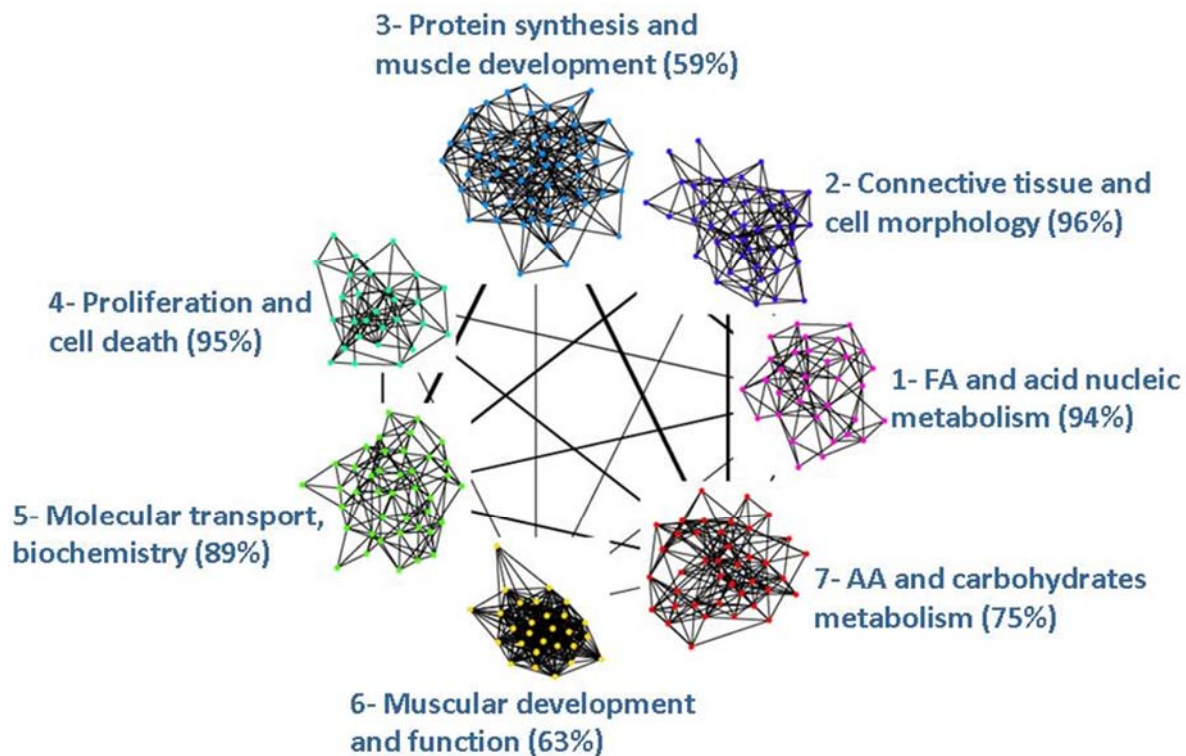


Figure 26 : Représentation simplifiée du réseau de co-expression des 272 gènes régulés par au moins un eQTL. La largeur de ligne entre les clusters est proportionnelle au nombre de liens entre les nœuds les clusters. L'information biologique associée provient de l'analyse d'enrichissement fonctionnel effectué avec le logiciel Ingenuity. Le pourcentage entre parenthèses correspond au nombre de gènes du cluster impliqués dans la fonction biologique indiquée.

Nous avons utilisé la corrélation partielle qui est la corrélation résiduelle entre deux gènes une fois prise en compte la corrélation avec tous les autres gènes. Ou dit autrement, la corrélation partielle permet d'évaluer la corrélation entre deux variables après avoir contrôlé l'effet perturbateur d'une ou de plusieurs autres variables¹. Nous avons utilisé une approche GGM (*Gaussian Graphical Model*). En misant sur les variations d'expression sans prendre en compte l'information eQTL, nous avons aussi recherché à identifier des groupes de gènes (communautés) et nous avons vérifié que ces communautés partageaient des fonctions communes.

Le résultat final (Figure 26) est le fruit de la collaboration avec deux statisticiennes dans le cadre de l'encadrement d'un stagiaire en master. Plusieurs modélisations ont été testées, plusieurs paramétrages évalués et la qualité de l'enrichissement fonctionnel systématiquement testé avec le logiciel WebGestalt (*WEB-based Gene Set Analysis Toolkit*). WebGestalt propose une visualisation hiérarchique de l'enrichissement en présentant les ontologies sous forme de DAG (*directed acyclic graph*). Les graphes d'enrichissement ont à un moment donné changé de « forme globale » montrant un véritable enrichissement pour chacune des communautés. A partir de ce moment-là, j'ai réalisé la même analyse d'enrichissement mais avec Ingenuity (IPA). Une moyenne de 80% des gènes impliqués dans les communautés et reconnus par le logiciel Ingenuity pour participer à des fonctions biologiques similaires. Dit autrement, pour une communauté de gènes co-exprimés, Ingenuity propose un réseau bibliographique.

¹ https://www.math.univ-toulouse.fr/~sdejean/Inra_Biostat/1_stat_explo.pdf

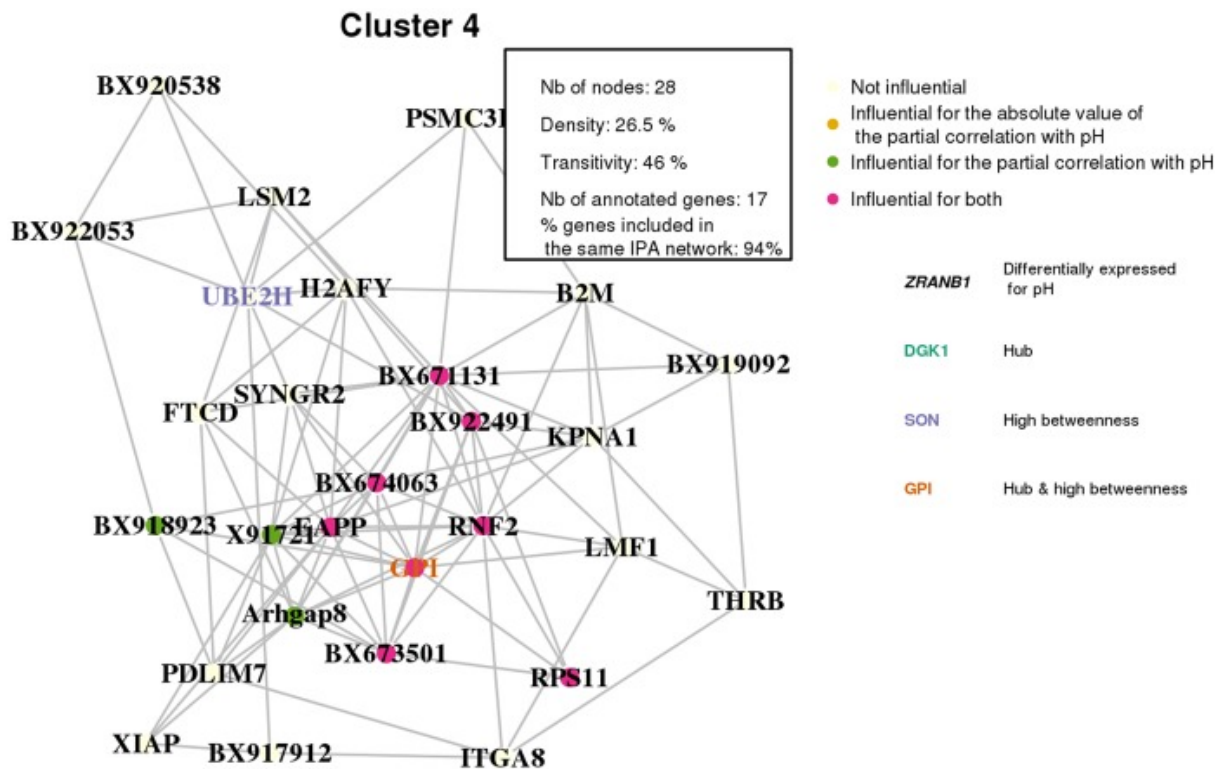


Figure 27 : Présentation détaillée du cluster 4.

Les nœuds (les gènes) qui ont une influence sur la corrélation avec le pH, ainsi que les nœuds qui sont importants pour la structure du graphe (hubs, forte centralité), sont mis en évidence. 94% des gènes du cluster sont reconnus par Ingenuity et identifiés comme étant impliqués dans un même réseau de gènes (réseau bibliographique) en lien avec la prolifération et la mort cellulaire.

Une attention particulière a été portée sur le cluster 4 pour lequel nous avons trouvé la plus forte corrélation avec un phénotype d'intérêt, le pH ultime (Villa-Vialaneix *et al.* 2013; Villa-Vialaneix *et al.* 2016). Parmi les 28 gènes du cluster 4, *GPI* (Glucose-6-Phosphate Isomerase ; enzyme de la glycolyse, Figure 28 ; $h^2 = 0.62$) a la plus forte valeur de centralité, ce qui rend ce gène particulièrement intéressant. La protéine GPI est connue pour être impliquée dans différentes voies métaboliques, surtout la glycolyse et la gluconogénèse, et connues pour être liées à la qualité de la viande. De plus, le gène *GPI* est localisé sur le chromosome 6 à la position de plusieurs QTL affectant le pH ultime dans le muscle de la longe. Dans ce jeu de données, l'expression de *GPI* est fortement corrélée au pHu. Il serait donc assez logique de proposer *GPI* en tant que gène candidat positionnel et fonctionnel (Li *et al.* 2010). Par contre, dans notre dispositif, il n'y a pas de *cis*-eQTL identifié et l'expression de *GPI* est régulée par deux QTL sur les chromosomes 5 et 8.

Une autre façon d'utiliser les données d'expression et de corrélation partielle est de rechercher une corrélation des gènes dans les clusters avec un phénotype d'intérêt, ici nous avons utilisé le pH dit ultime mesuré 24h après abattage. La question posée est de savoir si les nœuds reliés dans le réseau de co-expression ont tendance à être corrélés de la même manière avec le pH. Nathalie Vialaneix a utilisé l'Indice de Moran qui est une mesure de l'autocorrélation spatiale caractérisée par une corrélation entre les mesures géographiquement voisines d'un phénomène mesuré... Nous obtenons une mesure de corrélation par cluster indiquant que le cluster 4 est le plus lié au pHu. Nous obtenons également une valeur d'influence de chaque gène dans la corrélation avec le phénotype. Vingt gènes sont identifiés comme influençant et onze sont dans le cluster 4. Sur ces 20 gènes, 10 sont reconnus par IPA et inclus dans un même réseau bibliographique (Figure 28).

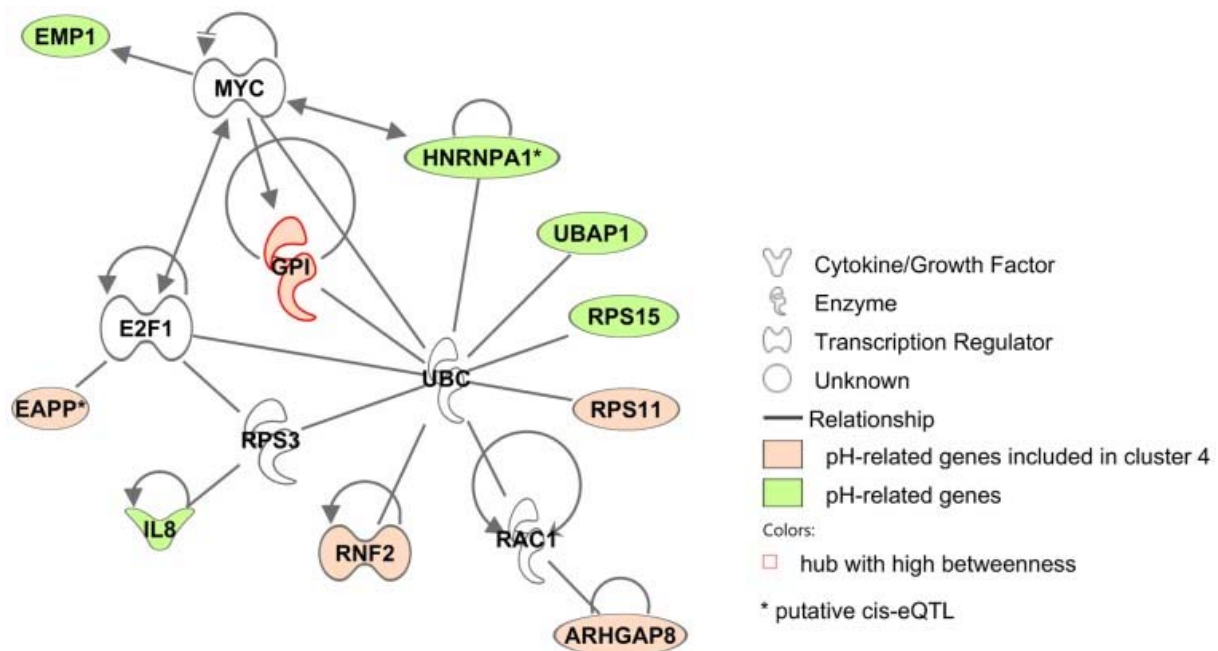


Figure 28 : Réseau bibliographique comprenant 10 gènes reconnus par IPA et reliés au pH ultime.

Dans le réseau IPA des gènes ayant une influence sur le pH, nous retrouvons le gène *GPI*. Deux autres gènes semblent intéressants, *EAPP* (2F Associated Phosphoprotein) et *HNRNPA1* (Heterogeneous Nuclear Ribonucleoprotein A). Pour ces deux gènes, de possibles *cis*-eQTL ont été détectés. *EAPP* est localisé sur le chromosome 7 et est régulé par un seul eQTL (en *cis*). La protéine codée par le gène *EAPP* est localisée dans le noyau tout au long du cycle cellulaire sauf pendant la mitose, il régule la transcription et la prolifération cellulaire. *HNRNPA1* est aussi un gène codant une protéine nucléaire qui intervient dans l'épissage alternatif des ARNm et leur transport. Les souris *hnRNPA1*^{-/-} meurent à un stade embryonnaire par défaut dans le développement musculaire (Liu et al. 2017). Ces deux gènes possiblement régulés chacun par un *cis*-eQTL sur des chromosomes différents est un exemple illustrant comment deux gènes candidats peuvent intervenir sur la régulation d'un caractère d'intérêt (épistasie). Par contre, des 20 gènes reliés au pH dans cette étude, seul *GPI* est connu pour avoir une fonction liée à la régulation de pH, plus précisément, *GPI* est impliqué dans la glycolyse et une glycolyse post mortem accélérée provoque une chute plus rapide du pH. Les anomalies de ce gène sont la cause, entre-autres, d'une mort néonatale immédiate.

Une étude similaire effectuée sur 326 porcs provenant de 4 familles F2 et analysée avec les mêmes 170 marqueurs microsatellites a identifié 9765 transcrits dont l'expression était significativement héritable, parmi lesquels 1174 présentaient une héritabilité supérieure à 0,5 (Cherel *et al.* 2012). Les auteurs ont également calculé des corrélations génétiques entre l'expression des transcrits et des phénotypes d'intérêt.

Ces approches génétiques (eQTL) et computationnelles (réseaux de co-expression) sont développées pour explorer la complexité de l'architecture génétique qui régule les caractères de qualité de viande. Les gènes identifiés comme ayant une « importance » particulière comme les *cis*-eQTL ainsi que les gènes ayant une forte centralité dans les réseaux de même que leur niveau d'expression est corrélé avec un phénotype d'intérêt, font de ces gènes de possibles biomarqueurs. C'est une façon de prioriser une série de gènes pour mener ensuite des études plus fines : suivre les variations de leurs expressions dans des dispositifs génétiques pour affiner l'identification de QTL et aller aussi jusqu'à l'identification de mutations causales.

Chez le porc, on note 35 publications dans Pubmed pour « pig eQTL » en juillet 2019. Ce qui est relativement peu depuis les années 2000, moment de l'essor de ces approches. Les deux tiers concernent la régulation génétique de l'expression dans le muscle. Plusieurs stratégies sont possibles dans les approches eQTL.

Elles peuvent être sans *a priori* et tester tous les gènes exprimés (Liaubet *et al.* 2011; Maroilley *et al.* 2017; Velez-Irizarry *et al.* 2019). Ce type d'approche génère beaucoup de données avec des problèmes méthodologiques (multiplicité des tests, importance déséquilibrée des *cis*- et *trans*-eQTL selon le type de dispositif expérimental...). Nous, nous avons choisi une approche sans *a priori* car nous cherchions plus à décortiquer des caractères complexes qui mettent en jeu des mécanismes cellulaires en interaction. Nous pensions que des gènes régulés génétiquement n'étaient pas forcément régulés ou différentiels avec le ou des caractères d'intérêt (et nous en avons plusieurs). Nous avons souhaité élaborer une autre stratégie comme présenté plus haut.

Une autre stratégie est souvent de présélectionner les gènes à soumettre à une analyse eQTL :

- Rechercher des eQTL pour des gènes impliqués dans des mécanismes biologiques précis. Par exemple, González-Prendes *et al.* (2017) ont sélectionné 63 gènes impliqués dans le métabolisme lipidique avant de rechercher des eQTL.
- Rechercher des eQTL pour des gènes corrélés au phénotype d'intérêt. Par exemple, Heidt *et al.* (2013) ont d'abord sélectionné 267 gènes significativement corrélés à la perte en eau du muscle post-mortem et impliqués dans des mécanismes biologiques similaires avant d'identifier des eQTL.
- Sélectionner des groupes d'animaux pour être extrêmes pour le caractère étudié. Exemple de l'odeur de verrat, pour ce caractère, les eQTLs ont été recherchés dans le foie et les testicules sur seulement 16 animaux par groupe. Les auteurs identifient ainsi 35 *cis*-acting eQTLs qu'ils proposent comme pouvant être inclus dans des stratégies de sélection d'animaux (Drag *et al.* 2018) pour être non porteur.
- Ne tester l'expression que des gènes localisés dans les régions QTL. Par exemple, González-Prendes *et al.* (2019) ont récemment choisi de tester 74 gènes localisés dans les régions QTL identifiées pour le taux de lipides intramusculaires.

Actuellement, les projets eQTL en cours d'étude chez le porc exploitent la puissance combinée de puces 60K SNP et de puces d'expression de 60K transcrits ou des données RNAseq. Nous avons de nombreuses données à analyser dans le cadre de projets ANR comme PigHeaT pour l'adaptation à la chaleur et SusOSstress pour l'étude des réponses de stress avec des approches permettant de combiner des données transcriptomes et génomiques. Les analyses combinées peinent à se réaliser, les données arrivant en fin de projet et les chercheurs impliqués mobilisés souvent sur de nouveaux projets. Les transcriptomes de ces projets (PigHeaT et SusOSstress) ont été acquis sur du sang total comme échantillon moins invasif que les échantillons de muscle. Donc ici encore la relation avec le phénotype d'intérêt pourrait paraître parfois éloignée mais il offre aussi la perspective de trouver et de tester des biomarqueurs.

Autre argument qui avait semblé intéressant, le transcriptome sanguin exprime environ 80% des gènes d'autres tissus (Mohr & Liew 2007). Avec les 6 transcriptomes du projet PORCINET dont le transcriptome sanguin, nous confirmons que nous avons effectivement 80% de gènes exprimés dans un tissu qui le sont aussi dans le sang. Par contre, les variations d'expression ne vont pas forcément dans le même sens entre le sang total et le tissu cible de l'expression du caractère étudié. Mais on peut quand même espérer identifier des eQTL cohérents ; nous nous intéressons à la régulation génétique et pas transcriptionnelle, les deux pouvant toutefois être liées... Des précisions sur les interactions entre *cis*- et *trans*-eQTL sont expliquées dans Le Mignon *et al.* (2010).

Peu ou pas de ces publications aboutissent à l'identification d'un eQTN, c'est-à-dire au polymorphisme régulant l'expression génique. L'article de Ma *et al.* (2014) combine une approche QTL et eQTL pour identifier une région QTL pour le potentiel glycolytique et ensuite utiliser l'approche eQTL pour identifier un gène régulé en *cis*. Ce gène est *PHKGI* (Phosphorylase Kinase Catalytic Subunit Gamma 1) qui est impliqué dans la régulation du métabolisme du glycogène. La mutation identifiée est responsable d'une accumulation de glycogène, l'enzyme codée étant peu ou pas fonctionnelle ; elle semble surtout présente chez les Duroc, absente chez les quelques sangliers analysés. Les auteurs considèrent cette mutation comme étant apparu avec la domestication et ayant un effet défavorable sur la qualité de la viande.

Globalement, les gènes identifiés pour être impliqués dans la qualité de la viande sont souvent des gènes codant des protéines ayant un rôle dans les métabolismes dont le fonctionnement doit être optimal à la naissance. Les gènes *PHGK1*, *GPI*, *COQ4* semblent être des cibles potentielles pour des analyses plus fines...

e. Recherche de gènes prédicteurs du phénotype.

Maintenant, je vais présenter des résultats non publiés, et même, non validés. Dans nos approches combinant des -omiques et des phénotypes, les objectifs sont souvent d'abord de décrire mais aussi de prédire un phénotype d'intérêt. Nous constatons que si nous consacrons beaucoup de temps à identifier les mécanismes biologiques qui sont impliqués dans l'expression des phénotypes d'intérêt, nous avons souvent des difficultés à réaliser le deuxième de nos objectifs, prédire. Nous nous retrouvons souvent hors projet pour rechercher des prédicteurs.

De plus, pour vraiment caractériser des prédicteurs, il faut les valider sur des dispositifs expérimentaux indépendants qu'on n'a généralement pas. Et pour les données d'expression, il faut valider par une autre technique moléculaire que celle qui a été utilisée pour identifier des prédicteurs ; qPCR vs microarray ou RNAseq. La qPCR demande du temps pour développer les amorces, les valider techniquement et les tester sur des jeux d'échantillons indépendants. Cette dernière partie est en cours de réalisation pour valider des prédicteurs de phénotypes mesurés à la naissance et en lien avec des questions de maturité, de croissance et de survie. Comme les résultats ne sont pas encore disponibles, je ne présenterai que la partie obtention de possibles prédicteurs pour le phénotype tendreté.

D'abord un point de sémantique. J'ai utilisé jusqu'à présent le terme de biomarqueur pour désigner des gènes dont la fonction connue et, si possible, dont les variations d'expression étaient bien significatives. Le terme de biomarqueur est plus générique, j'utiliserai le terme de prédicteur pour un gène ou autre (protéine, métabolite...) identifiés par des méthodes statistiques d'analyse discriminante qui vise à prédire l'appartenance à des groupes prédéfinis.

Pour la prédiction de la force de cisaillement comme phénotype du caractère tendreté, nous avons donc des difficultés à analyser l'ensemble de nos échantillons (Figure 18) d'une façon classique par comparaison d'extrêmes pour le phénotype. Nous avons obtenu très peu de gènes différentiels à un FDR de 20%. Néanmoins, j'avais persisté et essayé d'explorer de nouvelles façons d'analyser ces données.

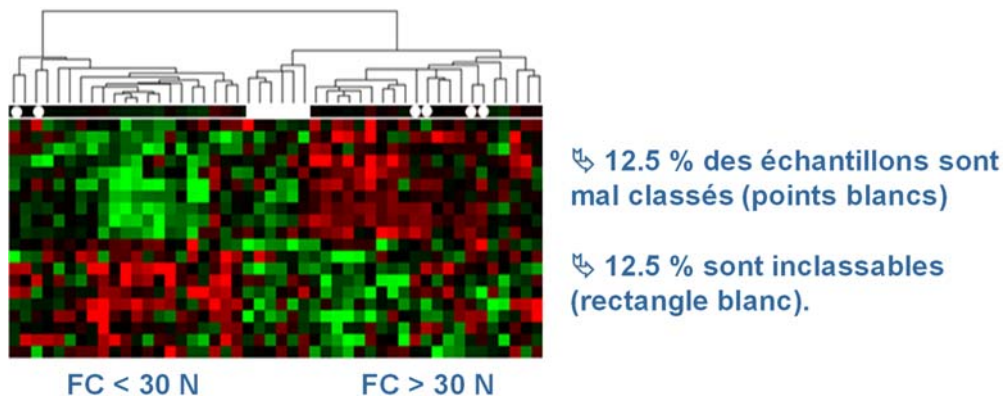


Figure 29 : Première recherche de possibles marqueurs de qualité par calcul d'un score discriminant sur l'ensemble des valeurs de force de cisaillement.

La première approche n'est pas nouvelle et est simplissime : il s'agit de calculer un score discriminant (Golub et al. 1999) pour chaque gène. Ensuite les gènes sont rangés par ordre décroissant des valeurs du score discriminant. Il suffit de prendre les 10 premiers et les 10 derniers (par exemple) du classement et d'utiliser les variations d'expression pour re-classer les échantillons. Sur la Figure 29, nous avons maintenant une signature moléculaire de 20 gènes et nous avons déjà une information intéressante, 25% des échantillons introduisent du bruit dans l'analyse, soit ils sont mal classés, soit ils sont inclassables.

Si nous avons restreint l'analyse des données aux 75% des échantillons qui correspondent à la signature moléculaire, nous aurions certainement réussi à identifier des gènes différentiels sur toute la gamme de force de cisaillement. Les informations moléculaires seraient alors tout à fait complémentaires de celles présentées précédemment.

Une approche similaire quoique plus sophistiquée a déjà été utilisée avec succès chez le poulet pour affiner la localisation de QTLs en regroupant d'abord les échantillons en quatre groupes selon leurs profils d'expression (Blum et al. 2011).

Ensuite, nous avons utilisé une méthode plus séduisante quoique au départ parfois jugée trop « boîte noire ». Cette méthode (ou famille de méthodes) est devenue très attractive. Il s'agit de la méthode des Forêts aléatoires (*Random Forest*, RF) développée par Breiman (2001). Cette méthode nous a déjà donné de très bons résultats pour classer des échantillons dans une autre thématique de l'équipe (Bonnet et al. 2008). Un autre avantage de la méthode des RF est de non seulement pouvoir discriminer des groupes mais aussi de pouvoir travailler en régression, adéquat pour un caractère quantitatif. Beaucoup de méthodes discriminantes ne fonctionnent qu'avec l'objectif de différencier des conditions. Les résultats de l'analyse par la méthode des RF sur 30 échantillons ont permis de définir un nombre minimal de six/neuf gènes permettant d'expliquer les valeurs de force de cisaillement (Figure 30). Cette étude a également fait partie du travail de Valérie Lobjois lors de son stage post-doctoral. Ces données n'ont pas été publiées car non validées expérimentalement. Et surtout valables que pour les valeurs de forces de cisaillement supérieures à 30N, donc d'un intérêt limité en l'état.

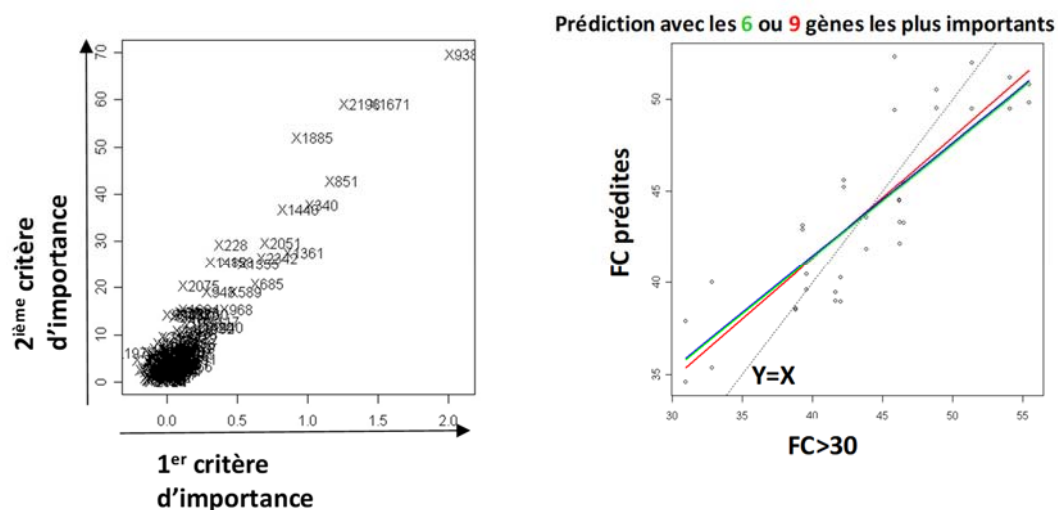


Figure 30 : Application de la méthode des Forêts aléatoires pour identifier les gènes les plus discriminants du phénotype force de cisaillement ($N > 30$).

La combinaison des approches statistiques de prédiction utilisant les données transcriptome et des données génétiques, nous a permis de proposer le transcrit meilleur prédicteur avec les RF comme candidat positionnel pour un QTL pour la tendreté (*taste panel tenderness*, (Malek et al. 2001; Rohrer et al. 2006)) et expressionnel. J'avais alors profité des expériences en cours de Laurence Drouilhet alors en thèse et de l'assistance de Julien Sarry, technicien dans notre équipe, pour tester si avec un fragment ADNc passé en SSCP (*single strand conformation polymorphism* ou polymorphisme de conformation des acides nucléiques simples brins), nous étions capables d'observer la présence d'un polymorphisme. Nous avons ainsi identifié un variant nucléotidique. Nous avons au final testé par séquençage 59 animaux au total : 20% de CC, 56% de CT, 24% de TT.

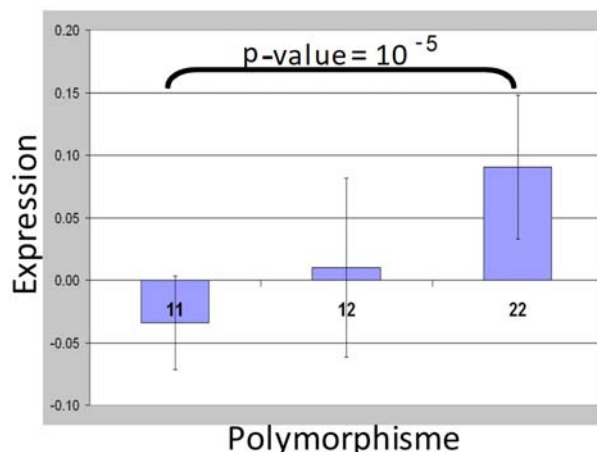


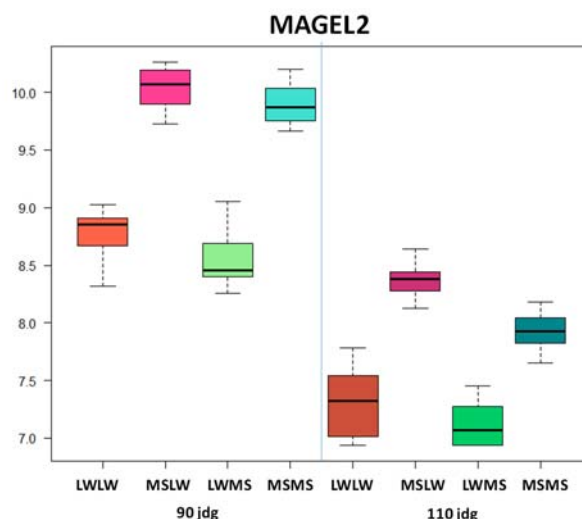
Figure 31 : Variations d'expression pour le meilleur prédicteur de la tendreté chez le porc selon le statut au polymorphisme identifié dans ce transcrit (données non publiées).

L'expression du prédicteur est effectivement différentielle (Figure 31) entre les échantillons selon le statut au polymorphisme et corrélé aux valeurs de forces de cisaillement si supérieure à 30 N.

Le meilleur prédicteur de la tendreté dans nos travaux est donc le gène noté X938 sur la figure 30. Un polymorphisme est présent dans le gène et l'expression du gène est différentielle selon les variants alléliques du gène. Ce gène est localisé dans une région QTL pour la tendreté ce qui en fait également un candidat positionnel.

Autre particularité de ce gène, son annotation a été très compliquée. Il se trouve sur le brin (-) de la séquence génomique avec plusieurs gènes plutôt très proches les uns des autres. D'autre part, c'est un gène sans intron, ce qui n'est pas si fréquent. Il n'est pas très simple de trouver des informations sur les gènes sans introns qui ne semblent pas particulièrement étudiés. Dans une revue publiée en 2015 (Hube & Francastel), les auteurs montrent comment les introns apportent une grande plasticité génomique aux génomes eucaryotes et en particulier aux mammifères. Les auteurs discutent aussi de la persistance de gènes sans intron chez les eucaryotes. Ils précisent que 1394 gènes codants (5,33% du total des séquences codantes annotées) et 3303 gènes non codants (38,06% du total des séquences non codantes annotées) sont constitués d'un seul exon. Bien que les gènes sans intron soient une caractéristique des procaryotes, un nombre significatif de ces gènes est toujours présent dans le génome humain. Il est intéressant de noter la forte proportion de ces gènes sans intron parmi les gènes non codants. D'un point de vue fonctionnel, la base de données SinEx (<http://www.sinex.cl/>) propose une analyse comparative des gènes sans intron dans les génomes de mammifères (Jorquera et al. 2016). Ainsi, les fonctions relatives à la structure de la chromatine, les mécanismes de transduction des signaux et la traduction sont significativement enrichies en gènes sans intron. Par contre, les fonctions relatives à la biogénèse des enveloppes, les métabolismes des protéines, nucléotides, lipides et métabolites secondaires sont significativement sous-représentées.

Dans les analyses d'expression réalisées sur le muscle fœtal, nous avons également identifié un gène sans intron par la méthode des RF capable de différencier les 4 génotypes aux deux stades. Il s'agit du gène *MAGEL2* connu chez l'Homme pour être responsable d'hypotonie néonatale, de retard de développement et d'obésité dans le syndrome de Prader-Willi (Tacer & Potts 2017). Attention, ce gène est décrit sans intron chez l'homme et la souris... et un intron vient d'apparaître (Ensembl release 97 - July 2019) dans ce gène chez le porc !? Guo et al. (2009) ont réalisé une étude ciblée sur le gène *MAGEL2* et ont caractérisé la structure moléculaire du gène ; ils n'ont pas trouvé d'intron...



*Figure 32 : Le gène **MAGEL2** identifié comme prédicteur du stade et du génotype dans le muscle fœtal identifié par la méthode des forêts aléatoires.*

L'intérêt de rechercher des prédicteurs exprimés dans le muscle est toutefois limité puisque relativement difficile d'accès du vivant de l'animal. Le muscle peut toutefois se collecter à l'abattoir ou par biopsie en élevage. La biopsie présente un risque de douleur pour l'animal qui rend le prélèvement difficile à prévoir sur de grands dispositifs d'animaux, difficulté à avoir l'autorisation du comité d'éthique... Aujourd'hui, la recherche de biomarqueurs/prédicteurs se focalise plus sur des tissus ou fluides plus « facilement » collectables (poils, salive, sang, fèces, urine...), donc probablement moins connectés aux phénotypes d'intérêt en terme de logique biologique. L'objectif est d'avoir une signature moléculaire du caractère d'intérêt.

D. Perspectives à court et moyen terme : Survie néonatale, intégration de données multi-omiques, multi tissus, prédiction à la naissance et intégration génétique.

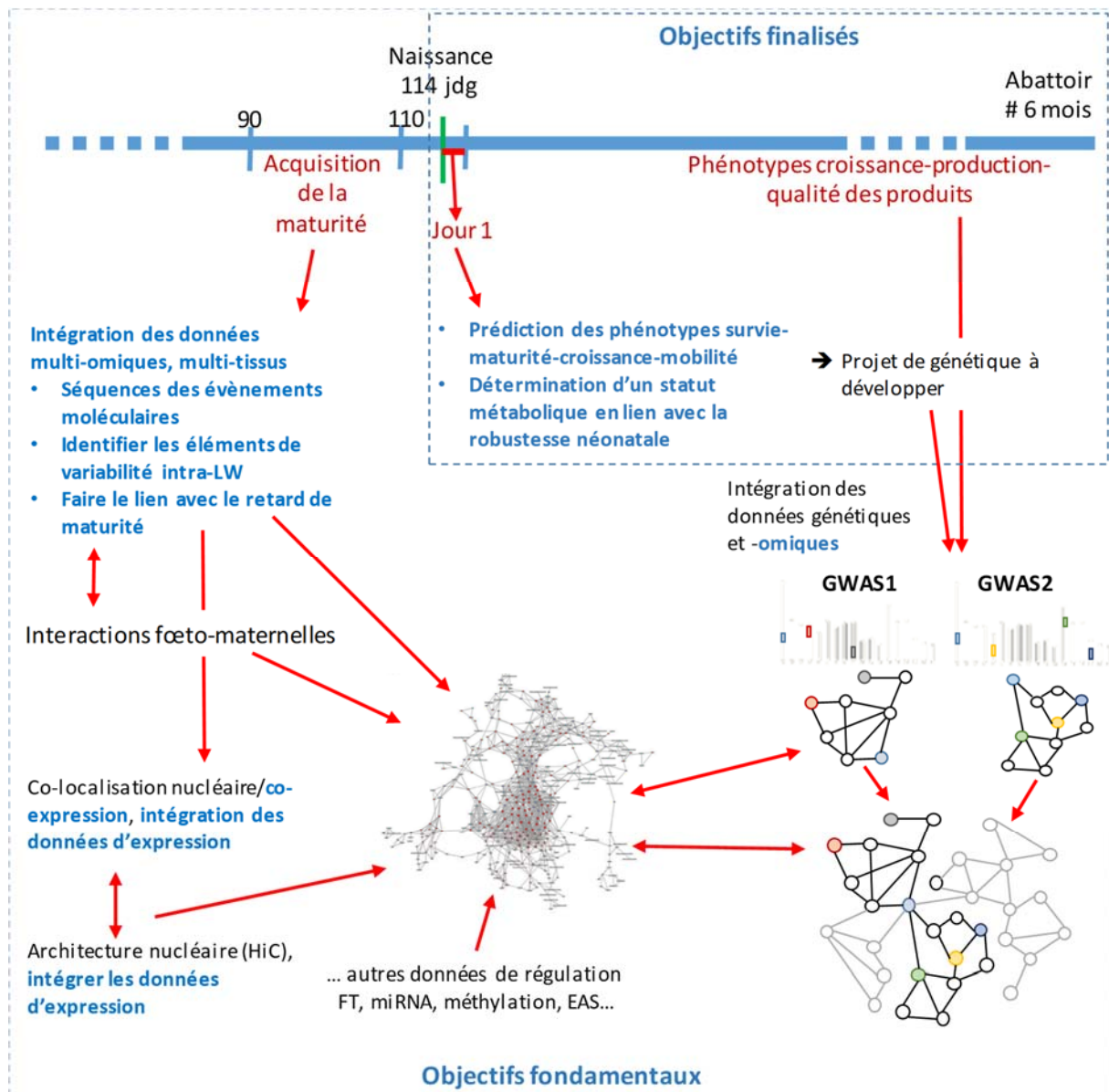


Figure 33 : Déterminisme précoce des phénotypes tardifs et intégration génétique-génomique.

En bleu, sont présentées les perspectives qui me sont propres et dans la limite des développements méthodologiques qui peuvent être nécessaires. En noir, les données qui sont travaillées d'abord par des collègues et collaborateurs.

Mes perspectives sont illustrées dans la Figure 33. Ce sont :

- Des perspectives à court terme pour répondre à des objectifs finalisés, prédiction de caractères en lien avec la survie néonatale et de la « qualité » de survie (maturité, croissance, mobilité).
- Des perspectives à moyen terme qui correspondent à tout le travail d'intégration des données du projet PORCINET. Il s'agira de mieux caractériser le développement en fin de gestation toujours en lien avec la survie néonatale. Il s'agira aussi d'évaluer comment ces données permette de caractériser la fin de gestation comme étape fondamentale du déterminisme de phénotypes plus tardifs.
- Des perspectives à long terme qui correspondent à l'intégration des différents types de données – omiques avec des données de génétique. Ce travail a déjà commencé avec les travaux antérieurs au

projet PORCINET ; il est en « toile de fond » actuellement et doit reprendre de l'importance dans les années qui viennent.

Donc pour parler de mes perspectives et de façon paradoxale, je vais replonger 10 ans en arrière. L'essentiel de mes travaux de recherche a porté, jusqu'à présent, sur le muscle en développement et à l'abattoir. Maintenant, je vais replacer les études réalisées sur le muscle fœtal dans un contexte plus général, celui du projet de recherche nommé PORCINET (ANR-09-GENM-005) que j'ai coordonné à partir de 2010.

Il y a dix ans maintenant, une « nouvelle » question est venue de la filière professionnelle. Il s'agissait de l'augmentation de la mortalité néonatale qui impactait négativement la filière aussi bien d'un point de vue économique qu'éthique. Le constat est que la sélection génétique pour plus de prolificité et plus de viande maigre aurait impacté négativement la capacité de survie des porcelets. Je passais alors d'une question spécifique limitée à un tissu à une problématique particulièrement complexe concernant l'animal entier.

Le projet PORCINET s'est construit sur quelques publications fondamentales pour la question posée mais aussi surtout sur deux thèses. L'une d'elle a été soutenue en 2001 aux Pays-Bas par Jascha Leenhouders, sa thèse « Biological Aspects of Genetic Differences in Piglet Survival » s'intéressait aux porcelets en fin de gestation, il avait réalisé des césariennes et les fœtus avaient été étudiés sur des aspects morphologiques avec quelques dosages fondamentaux comme ceux du glycogène et du cortisol. La deuxième est celle de Laurianne Canario, soutenue en 2006, le titre est « Aspects génétiques de la mortalité des porcelets à la naissance et en allaitement précoce : relations avec les aptitudes maternelles des truies et la vitalité des porcelets ». Laurianne s'est focalisée sur les porcelets entre la naissance et le sevrage ; elle a comparé différentes lignées génétiques (des lignées européennes et la lignée d'origine asiatique, Meishan ; Canario *et al.* 2006). Elle a aussi comparé le progrès génétique en étudiant la descendance de truies inséminées avec de la semence congelée en 1977 ou en 1998 (Canario *et al.* 2007).

Ce projet a donné lieu à une trentaine de communications mais seulement six publications. Quatre correspondent à la valorisation des primos-analyses et deux correspondent à des développements méthodologiques :

- Donc le transcriptome musculaire par Voillet *et al.*, (2014) et le protéome (Voillet *et al.* 2018).
- Le transcriptome du petit intestin (jejunum) par Yao *et al.* (2017). Ying Yao était une étudiante en thèse dans le laboratoire d'Isabelle Le Huërou-Luron (Unité ADNC, St Gilles). Yao *et al.* ont intégré de nombreuses caractéristiques morphologiques et enzymatiques de l'intestin ainsi que des données plasmatiques.
- Le protéome du tissu adipeux par Gondret *et al.* (2018).
- Une publication méthodologique relative aux analyses multi-omiques quand les données sont manquantes pour un échantillon dans un jeu de données afin de les intégrer à l'ensemble des données, par Valentin Voillet (Voillet *et al.* 2016),
- Et une deuxième publication méthodologique pour l'analyse des données métabolomiques, 1^{ère} publication de la thèse de Gaëlle Lefort (Lefort *et al.* 2019). La thèse de Gaëlle a débuté le 1^{er} septembre 2018.

Je vais donc présenter le dispositif expérimental complet, les données disponibles pour analyses et/ou valorisation ainsi que le potentiel lié à la biobanque constituée dans le cadre du projet PORCINET. Puis je présenterai le projet sur la prédiction des phénotypes à la naissance. Je finirai par quelques mots sur l'intégration génétique. Dans le texte, je mettrai en gras ce qui correspond effectivement à mes perspectives.

1. Intégration des données des fœtus en fin de gestation relatives à la survie néonatale

Au cours des dernières décennies, l'amélioration de la prolificité et de la composition corporelle (vers une viande plus maigre) a été le principal objectif en sélection porcine. Ce progrès génétique s'est accompagné d'une augmentation substantielle de la mortalité des porcelets avant sevrage. La période la plus critique est la période périnatale, essentiellement les 24 premières heures après la naissance. La mortalité précoce des porcelets dépend grandement de leur maturité à la naissance. L'objectif de ce projet est de tirer profit des connaissances actuelles sur deux races de porcs, Large White et Meishan, qui présentent des spécificités de maturité différentes. Pour atteindre cet objectif, nous avons développé une collaboration entre généticiens, physiologistes et biologistes moléculaires et de construire une importante base de données et de tissus fœtaux (bio-banque). La base de données contient des données phénotypiques et physiologiques, des profils métabolomiques, protéomiques et transcriptomiques obtenus sur différents fluides (3) et tissus (6). Nous avons également réalisé le transcriptome du sang total récupéré immédiatement à la naissance à partir du cordon ombilical (complément financier obtenue par L. Canario auprès du département de Génétique Animale de l'INRA).

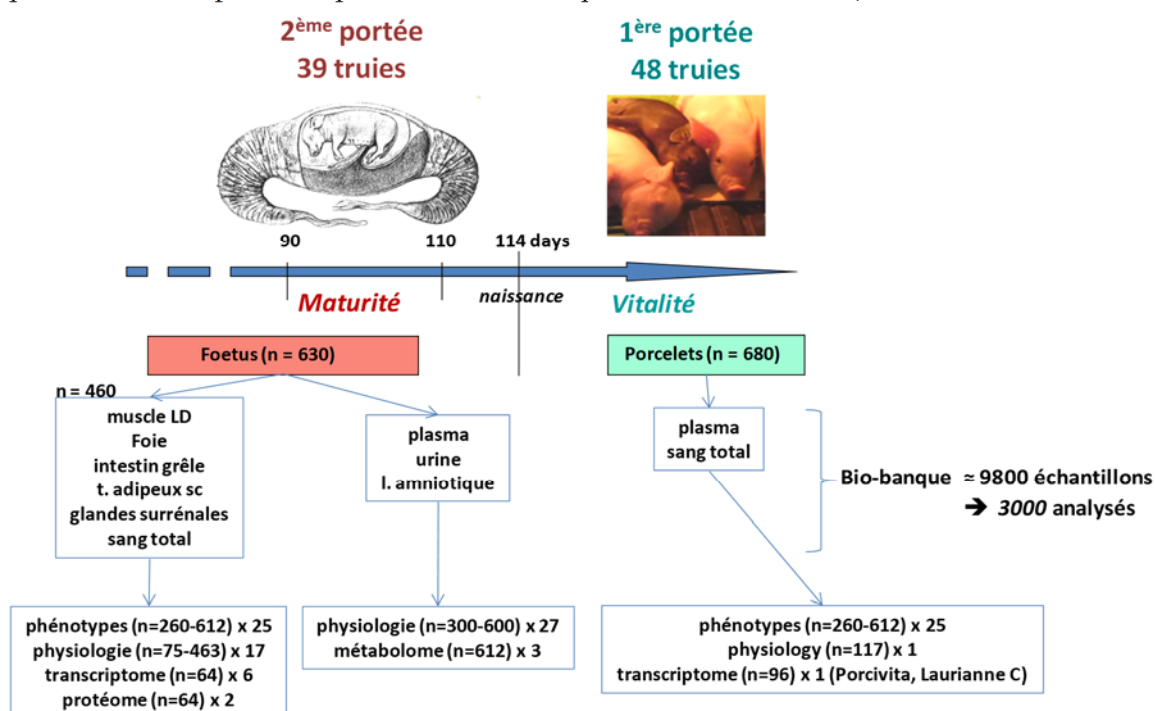


Figure 34 : Schéma expérimental du projet PORCINET.

Il s'agit d'un projet multi-tissus avec des prélèvements pour analyses multi-omiques et autres analyses (dosages plasmatiques et tissulaires, histologiques, enzymatiques, morphologiques).

Nous avons pour objectif de décrire la maturité du fœtus avant terme, un caractère complexe influençant la survie périnatale des porcelets, sur la base de deux races porcines et deux stades de la gestation. Par conséquent, le projet PORCINET a pour objectifs :

- de fournir des indicateurs comportementaux de la survie des porcelets,
- de préciser la physiologie et les bases moléculaires de la fin du développement fœtal,
- d'identifier de nouveaux biomarqueurs de maturité pré- et périnatale,
- d'étudier le contrôle génétique de ce processus en vue d'élaborer une stratégie pour des études de génétique quantitative et de sélection.

Ce projet impliquait trois laboratoires PEGASE, GenESI et GenPhySE. Les objectifs présentés ne font pas tous l'objet de mes perspectives, elles sont plutôt celles de ma collègue en génétique quantitative. Le dernier objectif, stratégie pour des analyses génétiques, pourra se faire surtout si nous arrivons à

fournir des biomarqueurs.

Le projet PORCINET est donc un projet intégratif qui va au-delà de l'étude du muscle. La mortalité néonatale correspond à une exagération d'un problème inhérent aux mammifères, l'adaptation à la vie extra-utérine ; capacité qui nécessite qu'à la naissance, tous les tissus et organes soient prêts à fonctionner et à assurer la respiration, l'alimentation, la motricité, la thermorégulation... Nous parlons de maturité définie comme l'état de plein développement permettant la survie à la naissance. Le dispositif permet d'avoir accès à trois stades de développement, 90 et 110 jours de gestation et la naissance (Figure 34).

Il reste à publier, en primo-analyses, les données phénotypiques des fœtus et des porcelets, les données physiologiques des fœtus, les données transcriptome du tissu adipeux sous-cutané, des surrénales et du foie, du transcriptome du sang total des fœtus et à la naissance.

La tissuthèque provenant des dissections et des prélèvements réalisés dans le cadre du projet PORCINET est aussi disponible pour d'autres études en cours ou à réaliser :

- Sang naissance pour le projet ANR PIPETTE, « Une approche -omique ciblée et intégrée pour comprendre le rôle des mécanismes d'empreinte génomique dans le déterminisme du poids de naissance chez le porc » (ANR-18-CE20-0018) coordonné par Julie Demars (équipe GenEPI de GenPhySE).
- Echantillons de placenta et d'endomètre pour le projet PENELOPE « Plasticité de l'Empreinte du placenta et de l'Endomètre en lien avec la maturité du Porcelet » (AAP département GA) d'Agnès Bonnet.
- Sont également disponibles : poumon, cœur, estomac, ovaires, cerveau, hypophyse, hypothalamus, plus tous les tissus et fluides déjà cités et donc partiellement utilisés.

Ce dispositif est utilisé pour des approches fondamentales de compréhension de la mise en place de la maturité essentielle à la survie à la naissance. **Après la publication des primo-analyses, il s'agira d'intégrer les données multi-omiques multi-tissus et multi-fluides avec également toutes les données phénotypiques biométriques et physiologiques.** Je vais préciser par la suite quelques aspects intégratifs encore assez génériques puis aussi quelques questions biologiques plus précises.

a. Trois jeux de données métabolomiques; plasma, urine et liquide amniotique.

Les premières analyses réalisées dans le cadre du projet PORCINET sont les analyses du métabolome des échantillons de plasma, d'urine et de liquide amniotique. Le premier avantage de l'analyse du métabolome a été de pouvoir analyser tous les échantillons (facilité de traitement des échantillons, coût minime de l'analyse RMN). Nous avons analysé les échantillons des 600 fœtus donc un total d'environ 1200 échantillons. Les analyses ont été réalisées par Cécile Canlet, responsable de la plateforme de RMN de Toulouse sous la direction d'Alain Paris aujourd'hui Professeur au Muséum National d'Histoire Naturelle de Paris. Ces données relativement faciles à acquérir, en comparaison des transcriptomes et protéomes, nous ont donné des résultats très importants (Figure 35) mais encore non publiés aujourd'hui. Importants car ils nous ont rapidement montré la puissance du dispositif, l'importance des deux stades choisis et l'impact des génotypes.

L'analyse du métabolome plasmatique a été effectuée sur les 612 fœtus par RMN 1H. 800 variables nommées *buckets* ont été obtenues après discrétisation des spectres bruts. Alain Paris a sélectionné 397 variables informatives par MDS (*Multidimensional Scaling*, positionnement multidimensionnel). La projection des individus sur la Figure 35 a été effectuée après O-PLS-DA (*Orthogonal Partial Least Squares*, projections orthogonales des moindres carrés partiels). Les résultats montrent que les profils plasmatiques séparent les fœtus sur le 1^{er} axe de l'O-PLS-DA selon les deux stades de gestation, 90 et 110 jours. Le 2^{ème} axe sépare les fœtus selon le type génétique de la mère (MS ou LW) à 90 jours mais pas à 110 jours de gestation. Ces résultats suggèrent que le rôle de l'environnement maternel diffère entre les deux stades.

A 90 jdg, les fœtus sont regroupés selon le génotype de leur mère, l'environnement maternel semble plus prédominer. A 110 jdg, les fœtus montrent une plus grande dispersion qui peut refléter une part de l'hétérogénéité de portée visible à la naissance et qui se mettrait donc en place en fin de gestation. La variabilité individuelle des fœtus s'exprimerait plus et donc moins sous le contrôle du génome maternel. **Autrement dit, à 110 jdg, nous avons trois génomes en interaction, les deux demi génomes parentaux s'exprimant chez le fœtus et le génome maternel constituant l'environnement. Ce constat nous incite à porter une attention particulière sur le déterminisme génétique de l'expression de cette variabilité.**

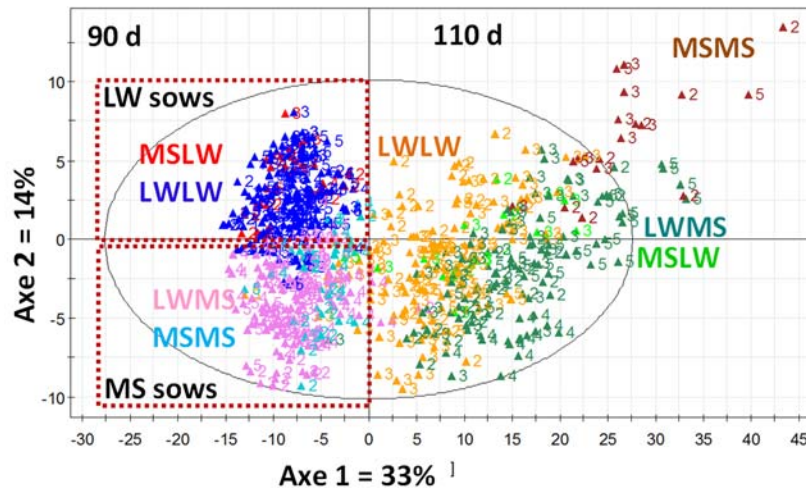


Figure 35 : Analyse discriminante par OPLS-DA des profils métaboliques plasmatiques par RMN¹H réalisée avec 397 variables identifiées comme informatives.

Les variables à ce stade sont anonymes, ce sont les *buckets*, entités représentant l'aire sous la courbe pour une portion définie du spectre RMN. L'analyse des données sous forme de *buckets* présente ici un fort pouvoir discriminant du caractère qui nous intéresse. En particulier, nous pouvons observer que l'axe 1 de la projection des individus montre pour les fœtus MS à 110 jdg sont les plus éloignés de l'axe alors que les LW connus pour être globalement moins mûres sont plus proches de l'axe voire certains se positionnent du côté des fœtus à 90 jdg. **Les coordonnées de la projection des individus sur l'axe 1 pourrait directement constituer un critère de maturité et servir dans les analyses intégratives avec les autres types de données disponibles.**

Il est donc possible d'identifier des *buckets* discriminants mais ils ne sont pas informatifs sur la biologie sous-jacente. Pour identifier les voies métaboliques, nous restons dépendants de l'expertise de l'ingénierie plateforme ou de bases de données de correspondance ; ce manque d'autonomie nous a incités à développer une méthode d'analyse automatique des spectres permettant de les traduire en données qualitatives et quantitatives. Cette question est d'autant plus importante que nous avons des données métabolomiques dans de nombreux projets en génétique donc obtenus sur des dispositifs important en nombre d'échantillons à traiter. **Ces données sont propices pour des intégrations génome-métabolome et identifier des mQTL en associant des analyses de scan du génome complet.**

Ces données RMN sont toujours en cours d'analyse. Une thèse a démarré en septembre 2018 avec Gaëlle Lefort. Sa thèse est financée par #DigiTag et les départements GA, SA et MIA. Les encadrants sont Nathalie Vialaneix (MIA), Rémi Servien (SA), Hélène Quesnel (PHASE) et moi-même. L'objectif est donc de développer la méthode ASICS, de la rendre disponible sous forme d'un package Bioconductor et d'analyser et d'intégrer les données de PORCINET et du projet SuBPig (prédiction à la naissance).

Gaëlle Lefort vient de publier la méthode ASICS, *Automatic Statistical Identification in Complex Spectra* (Lefort *et al.* 2019a). Le package R est disponible sous Bioconductor. Gaëlle a analysé les 1200 échantillons de plasma, d'urine et de liquide amniotique. La librairie de référence contient 190 métabolites, un total de 90 métabolites ont été identifiés, environ 60 par fluide et 39 en commun pour les 3 fluides. Les données sont en cours d'évaluation biologique. **Nous analyserons les trois**

métabolomes indépendamment afin d'étudier (comme pour le transcriptome et le protéome) les métabolites impactés par l'interaction génotype et âge gestationnel pour caractériser la maturité. Au-delà de la question de la maturité, des métabolomes obtenus sur des fœtus en fin de gestation sont en eux-mêmes des données originales.

Pour donner quelques exemples, une analyse par RMN du liquide amniotique a été réalisée chez des femmes enceintes avec des prélèvements au 1^{er}, 2^{ème} et 3^{ème} trimestre (Orczyk-Pawilowicz et al. 2016). Les auteurs observent eux aussi un switch dans les profils métaboliques entre le 2^{ème} et le 3^{ème} trimestre. Une revue résume pourquoi les approches métabolomiques chez la femme enceinte seraient un outil de diagnostic intéressant pour la santé de l'enfant (Palmas et al. 2016). Sinon des analyses métabolomiques sont réalisées sur du plasma récupéré à partir du cordon immédiatement après la naissance.

Nous rechercherons aussi à comprendre ce que ces trois fluides en lien les uns avec les autres nous apprennent sur les métabolismes mis en jeu.

Gaëlle Lefort utilisera aussi les différents transcriptomes disponibles pour aider à interpréter les résultats métabolomiques. En particulier, notre attention se portera sur la régulation des mécanismes régulant les stockages-déstockages des métabolismes énergétiques. Au final, nous aurons également les métabolomes plasmatiques en fin de gestation et à la naissance (projet SuBPig, données en cours d'acquisition). Nous pourrons évaluer la transition métabolique qui doit s'opérer et identifier les transitions défectueuses pour les fœtus / nouveau-nés.

b. Les six transcriptomes et deux protéomes

Les données des transcriptomes de cinq tissus (muscle, foie, tissu adipeux sous-cutané, surrénales ; n = 64 par tissu) et le transcriptome du sang total ont été acquises sur la plateforme GeT-TRiX (à Toxalim) par Yannick Lippi. Une vision globale des résultats transcriptomes est illustrée dans la Table 2. Elle donne le nombre de transcrits différemment exprimés entre les deux stades pour chaque analyse. Pour évaluer le nombre de gènes uniques différentiels, il faut savoir qu'il y a environ 50% de redondance sur la puce PORCINET. Ce sont les listes de transcrits différentiels avant sélection de modèles (Valentin avait développé jusqu'à 18 sous-modèles). Il est notable que l'impact du stade de gestation est aussi fort pour chacun des tissus que ce que nous avons décrit pour le muscle, soulignant que la transition entre 90 et 110 jours de gestation est bien une étape fondamentale dans le développement anténatal.

Tissu	Nombre de transcrits différentiels (seuil)	Nombre de protéines différentielles (seuil)	Principaux mécanismes biologiques identifiés
Glandes surrénales	14428 (FDR < 0,1%)	-	Développement, stéroïdogénèse
Muscle longissimus	12326 (Bonferroni < 1%)	435 (FDR < 1%)	Développement, adhésion cellulaire, métabolisme des carbohydrates, glyconéogénèse
Foie	11730 (FDR < 1%)	-	Phosphorylation oxydative, métabolisme des Acides aminés
TASC	17560 (FDR < 1%)	290 (p < 5%)	Métabolisme lipidique, métabolisme des carbohydrates, réorganisation du cytosquelette
Intestin grêle	9051 (FDR < 1%)	-	Système immunitaire, prolifération/apoptose, métabolismes mitochondrial et protéique

TASC : tissu adipeux sous-cutané ; FDR : 'False Discovery Rate' (taux de faux positifs)

Table 2 : Cinq tissus fœtaux analysés par étude du transcriptome et deux tissus par protéomique. Les nombres de DEG rapportés ici ont été obtenus avec le modèle complet développé par V. Voillet.

Une fois les primo analyses réalisées et valorisées, et avant d'effectuer des analyses intégratives, nous pourrons effectuer **une analyse comparative des cinq transcriptomes disponibles** analysés (plusieurs fois) avec exactement les mêmes procédures statistiques et avec la même puce Agilent 8 x 60K. En particulier, **cette comparaison pourra permettre de répondre à quelques questions génériques** : nombre de transcrits exprimés dans chacun des tissus et combien en commun, nombre de transcrits exprimés dans tous les tissus, nombre de transcrits exprimés que dans un seul tissu, nombre de transcrits

exprimés dans un tissu et aussi dans le sang total (recherche de biomarqueurs) ou dit autrement nombre de transcrits du sang total partagé avec chacun des autres tissus. Selon la question, des analyses fonctionnelles pourront être associées.

Parmi les nombreuses analyses statistiques réalisées, les toutes premières ont été effectuées par Yannick Lippi de la plateforme GeT-TriX. Il s'agissait d'une analyse comparative entre génotypes à stades fixes ou bien entre stades à génotypes fixés. Un exemple est donné dans la Figure 36. La comparaison des résultats peut donner une information intéressante. Il s'agit de la comparaison entre les génotypes deux à deux à 110 jdg dans les différents tissus. La plus grande différence est observée entre les deux génotypes extrêmes LW et MS pour tous les tissus (moins évident pour l'intestin). Les différences intra-génotype maternel sont assez homogènes. Par contre, le nombre de transcrits est très différent dans la comparaison intra-génotype paternel. Le nombre de transcrits différentiels est plus élevé quand le père est LW. Les fœtus qui ont une moitié de génome paternel LW sont plus variables selon l'environnement/l'autre moitié de génome maternel que si le père est MS.

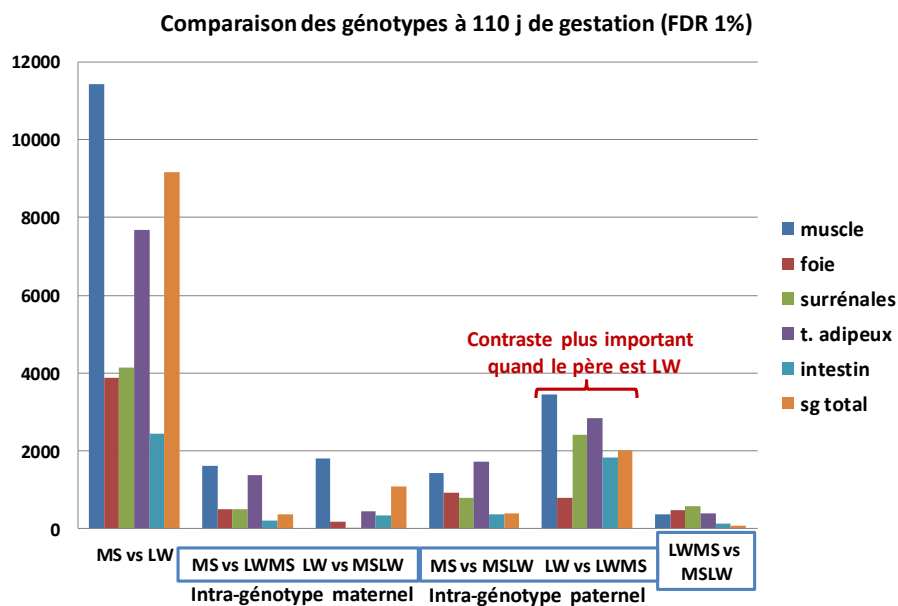


Figure 36 : Premières analyses statistiques des données des transcriptomes de cinq tissus et du sang total.

En débutant l'analyse intégrative des 6 transcriptomes, nous avons été vite confronté au problème des données manquantes. Plus précisément, chaque transcriptome était réalisé sur exactement les mêmes 64 échantillons. Pour des raisons techniques, un ou plusieurs échantillons peuvent manquer par transcriptome. Pour exemple, seulement 61 échantillons ont pu être utilisés pour le muscle. Dans les premières tentatives d'analyses intégratives, nous n'avions que 49 échantillons en commun. C'est dans ce cadre que Valentin Voillet a développé une méthode d'imputation de données appelée MI-MFA pour *Multiple Imputation-Multiple Factor Analysis* (package R missRows disponible sous Bioconductor (Voillet *et al.* 2016)). Il s'agit d'imputer des lignes manquantes (les individus d'un jeu de donnée – omique) dans des analyses multi-omiques.

Pour l'intégration des données -omiques, il y a deux grandes catégories de stratégies : des stratégies exploratoires (*data-driven*) ou non supervisées et des stratégies répondant à des questions biologiques précises.

Pour les stratégies exploratoires (sans *a priori* ou encore *data-driven*), il s'agira d'évaluer si chacun des transcriptomes globalement un comportement similaire à un ou plusieurs autres transcriptomes. Une analyse préliminaire avait été réalisée par Magali San Cristobal avec une approche MFA (*Multiple Factor Analysis*). Un exemple de résultat est donné dans la Figure 37 où les 8 conditions (4 génotypes et 2 âges) sont mieux séparées avec 15 gènes par transcriptome qu'avec chacun des transcriptomes individuels.

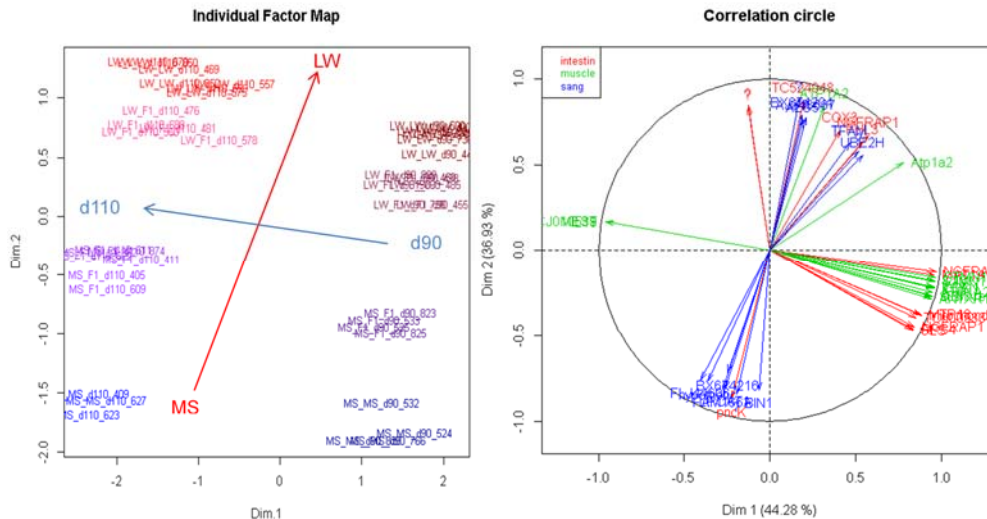


Figure 37 : Analyse MFA réalisée avec les 15 gènes les plus différentiels pour l'interaction génotype et âge gestationnel et pour les trois tissus : muscle, intestin et sang total.

L'Analyse Factorielle Multiple permet d'étudier plusieurs jeux de données complexes en équilibrant l'influence de chaque groupe dans l'analyse. Il est donc possible de comparer l'information apportée par chacun des groupes de variables. Dans cet exemple, nous pouvons évaluer la contribution statistique de chaque transcriptome dans la projection des individus mais aussi la contribution fonctionnelle de chaque transcriptome dans le déterminisme de la maturité à la naissance. Les 15 transcrits sanguins séparent mieux les génotypes (axe 2) et pas les stades. Ce sont des transcrits du muscle et de l'intestin qui différencient les stades 90 et 110 jdg.

Le transcriptome sanguin joue ici un rôle particulier. C'est le seul transcriptome accessible de façon la moins intrusive et délétère chez les nouveau-nés. Il sert donc ici plus à essayer d'évaluer comment il se positionne par rapport aux autres transcriptomes.

En plus de la MFA, nous utilisons l'environnement mixOmics (<http://mixomics.org/>) fruit du travail collaboratif entre Toulouse et l'Australie où est partie une ancienne étudiante en Master puis en thèse du laboratoire. Le package mixOmics propose différentes méthodes multivariées selon la question posée.

c. Données physiologiques

Ces données ont été acquises par Marie-Christine Père et Hélène Quesnel (PEGASE) ainsi que par Pierre Mormède (GenPhySE). L'ensemble des données ont été acquises sur les 600 fœtus ou bien sur un sous-ensemble d'environ 400 fœtus sélectionnés. Ces résultats n'ont pas été publiés dans leur ensemble. Ces résultats ont été analysés et certaines données intégrées aux données -omiques c'est-à-dire pour 48 ou 64 fœtus. J'insisterais sur quelques mesures physiologiques qui sont des références dans le cadre de l'étude de la survie néonatale.

Une des mesures physiologiques les plus importantes pour la survie néonatale est la teneur en glycogène des tissus de réserve que sont le muscle et le foie. Comme déjà présenté pour le muscle, Figure 5 pour 64 fœtus et Figure 38 pour 459 fœtus, le glycogène s'accumule en fin de gestation de façon très importante au-delà des quantités observées chez l'adulte au moins pour le muscle. La découverte que le muscle et le foie des nouveau-nés sont particulièrement riches en glycogène n'est absolument pas nouveau. Il y a plus de 150 ans, Claude Bernard lui-même publiait sa découverte de la molécule de glycogène (1857) et l'étudiait, en particulier, dans les heures suivant la naissance chez des mammifères. Une rétrospective de ses travaux a été publiée 100 ans plus tard en 1957 (Young 1957). En 1986, une revue récapitule toutes les informations sur les mobilisations des réserves énergétiques à la naissance (Figure 39) chez l'homme, le mouton et le porc (Mellor & Cockburn 1986). Les réserves lipidiques étant très faibles à la naissance chez le porc, le glycogène est particulièrement important chez les porcelets.

Toutefois, dans notre dispositif expérimental, les fœtus Meishan ont un tissu adipeux sous-cutané niveau dorsal plus riche en triglycérides que celui des Large White (Gondret *et al.* 2018). La Figure 38 montre que le glycogène double entre 90 et 110 jdg dans le muscle, un résultat similaire a été obtenu pour le glycogène hépatique (MC Père, PEGASE, données non publiées). Comme indiqué sur la Figure 39, ce stock peut être épuisé dans les 48 h suivant la naissance. Ce résultat fait échos à ceux qui ont motivé le développement du projet PORCINET. Initialement des travaux réalisés chez le porc en fin de gestation et à la naissance avaient montré que le stockage du glycogène était un paramètre déterminant du potentiel de survie (van der Lende *et al.* 2001; Leenhouders *et al.* 2002; Mota-Rojas *et al.* 2011).

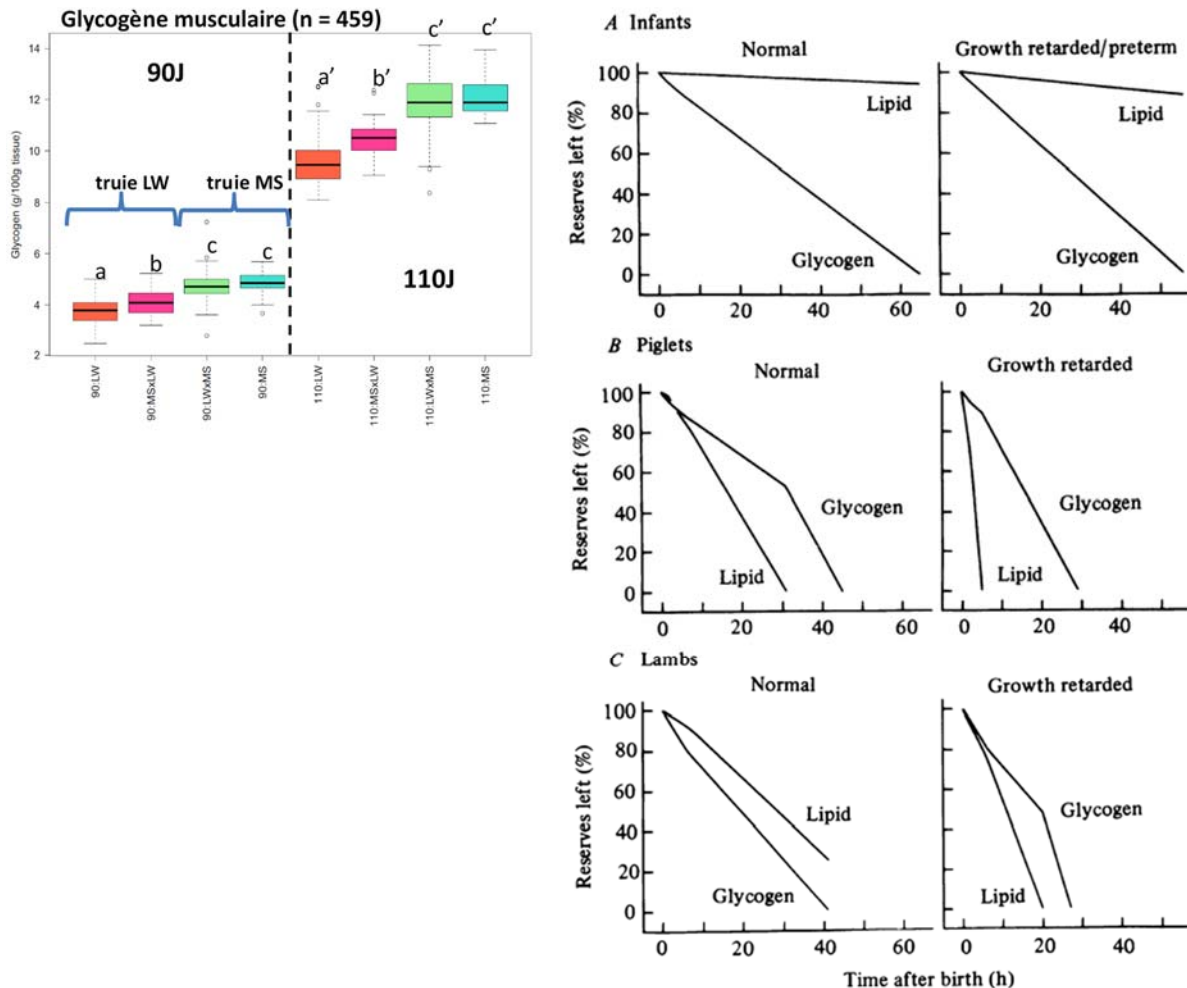


Figure 38 : A gauche. Glycogène musculaire chez 459 fœtus.

La quantité de glycogène double (au moins) entre 90 et 110 jdg. Le glycogène musculaire est plus important chez les fœtus de mère MS que LW, il est stable intra-MS et plus élevé intra-LW si le père est MS.

Figure 39 : A droite. Déplétion des réserves énergétiques dans les heures qui suivent la naissance.

Reprise d'une figure de la revue de Mellor and Cockburn publiée en 1986 reprenant les résultats obtenus chez plusieurs mammifères (Mellor & Cockburn 1986).

Canario *et al.* (2007) avaient également observé qu'à la naissance, des porcelets issus de truies LW inséminées avec de la semence LW récoltée en 1998 avaient moins de glycogène hépatique que les truies LW inséminées avec de la semence de 1977. Et donc à la naissance, ce stock énergétique est le premier à être épuisé. **Il s'agira, dans les approches intégratives, de comprendre quels sont les facteurs favorables à ce stockage en combinant les transcriptomes du muscle et du foie avec le protéome du muscle, les données physiologiques plasmatiques cliniques (glucose, fructose) ainsi que le métabolome plasmatique. A un autre niveau d'intégration et vu le rôle que joue le glycogène dans des caractères de qualité de la viande, il est possible que la sélection génétique pour une viande dont le potentiel glycolytique est moins néfaste ait eu un rôle préjudiciable sur la capacité du muscle en fin de gestation à stocker et à mobiliser ses réserves à la naissance de façon efficace. Il**

sera intéressant de focaliser sur la régulation des voies métaboliques impliquées et son impact autour de la naissance et sur la qualité de la viande.

L'autre mesure, endocrinienne, importante est celle du cortisol plasmatique. Comme pour le glycogène et même en amont du glycogène d'un point de vue fonctionnel, le cortisol joue un rôle fondamental en fin de gestation pour son rôle dans la maturation des tissus et organes périphériques. Ici, l'axe corticotrope montre un rôle autre que celui qui nous intéresse par ailleurs, les réponses de stress. Ces travaux sont menés en collaboration avec Pierre Mormède spécialiste du rôle de l'axe corticotrope dans les réponses de stress et Elena Terenina plus spécialiste des aspects moléculaires des réponses de stress et du transcriptome des glandes surrénales en particulier. Avec Elena Terenina, nous avons encadré un stagiaire en Master, pour l'analyse des données transcriptome. Les résultats ne sont pas encore publiés. Comme indiqué dans la Table 2, les fonctions différentiellement régulés concernent le développement et la stéroïdogénèse. La deuxième fonction en lien avec les régulations endocrines intéresse plus Elena Terenina. La première concerne directement le développement des surrénales et plus précisément un retard et une grande variabilité d'expression de plusieurs gènes à 90 jdg et presque uniquement chez les fœtus LW. Ces gènes codent des protéines du cytosquelette et leurs expressions diminuent drastiquement à 90 jdg, sauf chez la LW. Ceci semble être en relation avec une modification morphologique majeure dont nous trouvons peu de possibles explications dans la littérature. Toutefois, une telle modification a effectivement été observée indépendamment de ce projet par nos collègues Anne Tarrade et Pascale Chavatte-Palmer de la BDR (Biologie du Développement et Reproduction, INRA, Jouy-en-Josas). Ma collègue Agnès Bonnet a développé une collaboration avec nos collègues de Jouy pour mieux caractériser les mécanismes morphologiques au niveau moléculaire et histologique. Ces résultats sont en cours de valorisation.

Le rôle dans la maturation des tissus nécessite donc que les glandes surrénales soient prêtes dès 90 jdg dans leur rôle de production du cortisol. Le cortisol augmente en fin de gestation (Leenhouwers et al. 2002 ; nos données). Il semble donc que le retard de développement décrit pour le muscle en fin de gestation serait déjà observable dans les surrénales chez les LW à 90 jdg.

Une des hypothèses de travail de notre équipe sur la question de la robustesse prend ici tout son sens. Mormède et al. (2011) expliquent dans leur revue l'impact possible d'une sélection pour des caractères de production sur des caractères de robustesse comme ceux mis en jeu dans la survie néonatale. La sélection pour des caractères de production aurait affaibli la réponse corticotrope aux différents stress. Or le cortisol joue un rôle important en fin de gestation, sa concentration plasmatique doit normalement augmenter. Il est impliqué dans la maturation fœtale et donc des organes et tissus pour les préparer à la naissance. Ce mécanisme est bien connu et des corticoïdes sont administrés à la mère (chez les humains) ou bien au nouveau-né prématuré pour activer la maturation des poumons.

L'objectif en terme de biologie intégrative sera de modéliser quels mécanismes moléculaires et pour quels effets sur les tissus et organes sont mis en jeu pour coordonner et réguler la maturation. Il s'agira d'identifier des leviers pour optimiser le statut métabolique du nouveau-né, pour une meilleure survie. Pour la génétique, il s'agit d'identifier plus finement les interactions entre fonctions de production qui conditionnent plus la physiologie musculaire (par exemple) et les fonctions de robustesse comme la survie néonatale. Pour compléter l'étude de la régulation endocrinienne, des analyses de l'hypothalamus et l'hypophyse permettront de compléter la cascade des événements moléculaires. Il pourra s'agir d'études d'expression mais aussi d'étudier les mécanismes épigénétiques. Plusieurs gènes impliqués dans l'axe corticotrope sont déjà caractérisés pour avoir des niveaux de méthylation différents selon le niveau de stress de la mère (Howland et al. 2017). Ces mécanismes affectent durablement le devenir des individus, leur vie future et aussi leur descendance.

d. Données morphologiques fœtales et chez le nouveau-né

La caractérisation des fœtus en fin de gestation (environ 600 fœtus) et à la naissance (environ 600 nouveau-nés) a été réalisée par Laurianne Canario (équipe MODGEN, GenPhySE). Il s'agit de l'analyse des caractéristiques morphologiques qui peuvent influencer la survie, la vitalité et la croissance. Que ce soit chez le fœtus ou le nouveau-né, le poids, la longueur cranio-caudale et la longueur du fémur

ont été mesurés. Des mesures allométriques ont été réalisées sur les fœtus : le poids de la tête et du cerveau, les poids du cœur, du foie, poumons, estomac, reins. Les ratios entre les différents organes ont été calculés, en particulier par rapport au poids du cerveau. Le cerveau se développe en premier et donc des variations du ratio peuvent être mises en relation avec des retards de croissance. D'autres paramètres tels que l'Indice de Masse Corporel (IMC) ou l'Index Pondéral (IP) ont aussi été calculés.

Chez les nouveau-nés, d'autres critères sont évalués en lien avec la vitalité. La mobilité dans la boîte de pesée est évaluée par un score ainsi que la respiration et les vocalisations. Des pesées ont aussi été réalisées à 24h, à 3 jours et à 7 jours. Un certain nombre d'enclos étaient équipés de caméras vidéo. Les truies et les portées ont été filmées de la naissance jusqu'au sevrage.

Pour l'instant, nous n'avons communiqué que quelques résultats aux Journées de la Recherche Porcine (Canario et al. 2014) sur un sous-ensemble de 19 portées correspondant à 312 fœtus. Les informations sorties de ces analyses sont déjà utilisées par Laurianne Canario et moi-même dans de nouveaux projets dont un sur la validation de possibles biomarqueurs de la survie et la maturité en cours de réalisation.

Une première intégration de données a été réalisée par Laurianne Canario par sparse PLS entre les données phénotypiques et le transcriptome du sang à la naissance (Figure 43). L'objectif était d'identifier des gènes covariants avec les phénotypes d'intérêt. J'ai pour ma part plutôt travaillé sur la prédiction de la mortalité (Figure 42). Voir le chapitre 3 des perspectives.

En termes d'intégration, l'objectif est plus une question de recherche de biomarqueurs/prédicteurs que d'expliquer les processus biologiques sous-jacents. Les indicateurs recherchés peuvent être multiples ; morphologiques et moléculaires. Il s'agira de trouver la meilleure combinaison pour évaluer les différents caractères en lien avec la survie néonatale et d'autres phénotypes. Sinon l'ensemble des phénotypes fœtaux constitue un jeu de données complémentaires à utiliser dans les approches intégratives. L'objectif de cette tâche est d'abord de fournir à Laurianne Canario des indicateurs qui pourrait être utilisés dans un projet de génétique quantitative, projet à développer.

Globalement ces résultats ont montré que la production de fœtus purs et croisés au sein d'un même environnement utérin permettait de dissocier les influences génétiques maternelles et fœtales. Canario et al. (2014) précisent que même au sein d'un environnement plus homogène qu'offre les truies MS par rapport aux truies LW, les fœtus de types génétiques différents ont un développement différent. **Nous soulignons ainsi le rôle important de la génétique du porcelet dans son développement en fin de gestation.**

2. Etude des interactions fœto-maternelles

Au cours des séries de prélèvements réalisés dans le cadre du projet PORCINET, les placentas ont été phénotypés (poids, densité des microcotylédons (ou aréoles), vascularité) et les tissus adjacents du chorion (génome fœtal) et de l'endomètre (génome maternel) ont été prélevés et congelés séparément. Nous avons des échantillons pour environ 400 fœtus. Les études préliminaires réalisées sur ce sujet ont été confiées à une collègue de l'équipe, Agnès Bonnet (Ingénieure de Recherche) et effectuée en collaboration avec A. Tarrade (UMR BDR). Malheureusement nous rencontrons de grandes difficultés pour financer des études plus poussées pour cette question de recherche très fondamentale.

La sélection génétique pour la survie des porcelets a mis en évidence une corrélation positive entre la survie des porcelets et le rapport poids fœtal/poids placentaire ; ce rapport est principalement utilisé comme mesure indirecte de l'efficacité du placenta (Leenhouwers *et al.* 2002). Inversement, la sélection génétique pour l'efficacité placentaire n'a pas amélioré le taux de survie (Mesa et al. 2012). Des comparaisons entre une race moderne (Yorkshire, Y) et une race rustique (Meishan, MS) ont montré que le génome du fœtus contrôlait l'efficacité et la vascularisation placentaire en fin de gestation (Biensen et al. 1999). Un bon équilibre entre la demande en éléments nutritifs pour une croissance fœtale optimale et les besoins en éléments nutritifs de la mère est nécessaire pour soutenir la grossesse, l'allaitement et les portées ultérieures. Le placenta joue un rôle central dans ce dialogue croisé. Il régule l'allocation fœto-maternelle des ressources en adaptant son phénotype (taille, morphologie et vascularisation) et sa fonction (échanges transplacentaires, synthèse nutritionnelle et métabolisme, régulation des molécules,

etc...). Par conséquent, le placenta est un acteur central de la programmation fœtale qui pourrait avoir des effets immédiats et à long terme sur la croissance et la maturité du fœtus, la santé après la naissance et le phénotype adulte faisant référence au concept de la DOHAD (*Developmental Origins of Health and Disease* ; Tarrade et al. 2015). La physiologie et la nutrition maternelles pendant la grossesse sont également intimement liées à la croissance fœtale, car les nutriments doivent être mobilisés et livrés au placenta, puis au fœtus. La croissance fœtale et placentaire est maximale en fin de gestation et entraîne un fardeau nutritionnel pour la mère. Enfin, Workalemahu et al. (2018) suggèrent que les facteurs environnementaux ont une influence plus forte sur la croissance en début de gestation, mais que les influences génétiques prédominent en fin de gestation. Nous-même, nous démontrons qu'en fin de gestation (Figure 4, 5, 38 ; (Voillet *et al.* 2014; Voillet *et al.* 2018)), que ce soit avec les données du transcriptome musculaire ou du métabolome plasmatique, les fœtus plus modernes (LW) présentent un certain retard de développement en fin de gestation, l'expression phénotypique semble moins sous le contrôle de l'environnement maternel (utérin). Jusqu'à présent, la survie des porcelets était considérée comme un trait de sélection génétique influencé par la mère et la contribution du génome des porcelets semble sous-estimée.

Nos précédents résultats sur l'analyse du transcriptome du muscle fœtal ont montré que l'expression de certains gènes pouvait être influencée préférentiellement par le génome paternel (366 gènes) ou bien par le génome maternel (106 gènes). Ces gènes ont été sélectionnés dans un modèle où ils sont en interaction avec l'âge gestationnel puisqu'on s'intéresse d'abord à la maturité (sous-ensemble des gènes différentiels entre les deux stades et différents entre les génotypes). Des informations sur les gènes soumis à empreinte chez le porc et dans le placenta en particulier peuvent être trouvées dans la publication de Bischof et al. (2009). Les gènes que nous avons identifiés peuvent être des gènes effectivement soumis à empreinte ou bien cibles de ces gènes, ce qui peut expliquer les variations observées que nous appelons d'expression préférentielle paternelle ou maternelle.

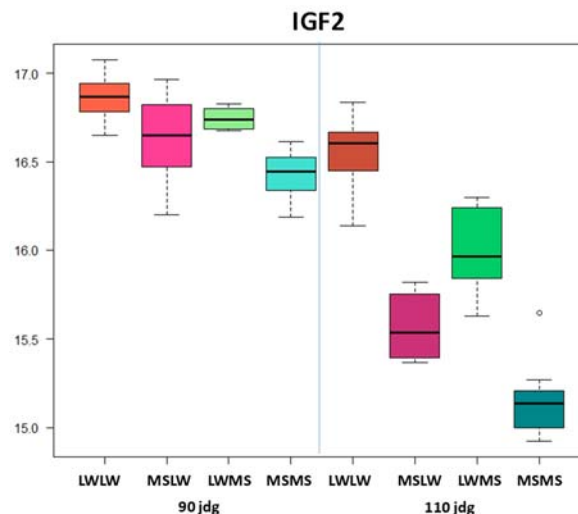


Figure 40 : Exemple du gène IGF2 d'expression préférentiellement paternelle dans le muscle fœtal.

Les Figures 32 et 40 illustrent l'expression préférentiellement influencée par le génome paternel pour deux gènes, *MAGEL2* et *IGF2*. *MAGEL2* (MAGE Family Member L2), ce gène sans intron dont j'ai déjà parlé, n'a pas une fonction complètement élucidée. Le gène *MAGEL2* (Figure 32) a été identifié comme étant impliqué dans le syndrome Prader-Willi où on observe une hypotonie néonatale, du retard de développement et de l'obésité (Tacer & Potts 2017). Une étude réalisée chez le porc (Guo et al. 2012) démontre l'expression monoallélique (paternelle) ; les auteurs ont identifié un polymorphisme dans ce gène associé à des caractéristiques de la carcasse. Guo et al. (2012) ont également recherché l'expression de *MAGEL2* dans différents tissus et ils ont trouvé une expression paternelle, entre autres, dans le muscle et le placenta. Dans une autre étude sur *MAGEL2* dans le placenta (Jiang et al. 2014), des précisions sont données sur les modalités de régulation de ce gène et une association est également identifiée entre un polymorphisme et l'efficacité placentaire. Je ne reparle pas de *IGF2* (Figure 40) qui a été au centre d'une

de nos études (Marti-Marimon *et al.* 2018) décrite dans la partie Bilan (Chapitre C1d). *IGF2* est peut-être un des gènes soumis à empreinte les plus étudiés pour son rôle dans le développement du muscle (Jeon *et al.* 1999; Nezer *et al.* 1999; Van Laere *et al.* 2003; Marti-Marimon *et al.* 2018).

Agnès bonnet a réalisé une étude préliminaire d'expression de 40 gènes dans le placenta (le chorion, étude d'expression allèle spécifique) ainsi que dans l'endomètre (beaucoup moins étudié). Pour rappel, l'endomètre correspond à un seul génome, celui de la mère, en face duquel se trouvent les différents génomes des différents fœtus. Nos résultats préliminaires ont montré que l'expression de certains gènes dans l'endomètre est influencée par le génotype ou le sexe du fœtus à 90 jours de gestation, avec une sensibilité plus importante de l'endomètre LW au fœtus juxtaposé par rapport à un endomètre MS. Le fœtus pourrait donc influencer lui-même les capacités utérines et donc avoir un rôle dans les échanges fœto-maternels nécessaires à sa propre croissance (Lunde *et al.* 2007).

Plusieurs gènes soumis à empreinte ont un rôle tout au long de la gestation. Les gènes à expression maternelle seraient souvent des gènes « d'épargne », ayant un effet négatif sur la croissance, permettant à la mère de survivre à la gestation et de pouvoir mener d'autres gestations. Les gènes à expression d'origine paternelle seraient plus fréquemment des gènes de croissance, favorisant la croissance de la descendance directe.

Il nous semble donc qu'il serait très important de caractériser finement les interactions fœto-maternelles afin d'identifier des leviers qui pourrait être optimisés pour favoriser la maturité et la survie des porcelets à la naissance. Ces résultats suggèrent l'importance d'acquérir des connaissances sur les contributions des gènes fœtaux et maternels au cours du processus de maturité (à partir de 90 jours de gestation). Nous souhaitons également avoir une approche métabolomique des deux tissus adjacents pour ajouter un niveau d'information sur la qualité des échanges mère-fœtus et l'impact sur la répartition des ressources. Pour finir, nous souhaiterions réaliser l'intégration de l'ensemble des données, expression de gènes-métabolome côté maternel et expression de gènes-métabolome côté fœtal.

3. Prédiction de la survie et de la qualité de survie à la naissance.

Pour la première fois, j'ai pu obtenir le financement d'une phase expérimentale post-génomique pour valider de possibles prédicteurs. Comme je l'ai précisé précédemment, pour pouvoir prétendre avoir des prédicteurs, il faut les valider sur des dispositifs indépendants. Je vais pouvoir, dans l'année qui vient, savoir si effectivement nous pouvons proposer des prédicteurs de phénotypes quantitatifs ou qualitatifs. Nous avons pu effectuer environ 600 prélèvements et mesures à la naissance sur 590 nouveau-nés. La Figure 41 donne un exemple de l'hétérogénéité qui peut être observé à la naissance.

Pour réaliser le projet, il a fallu un gros investissement en temps de toute mon équipe ; ingénieures, techniciennes et stagiaires. Nous n'avons pas encore toutes les données, nous avons commencé à explorer les premiers résultats. J'expliquerai donc la genèse, l'évolution de ma réflexion sur la procédure et les résultats espérés.

Jusqu'à présent, nous avons donc pu aborder les questions de prédiction dans nos projets surtout pour la partie identification. Ce que nous appelions validation était l'étape de test des variables sélectionnées sur les mêmes échantillons à partir desquels ils avaient été identifiés. En général, les méthodes déployées comprenaient un partage du jeu de données en, par exemple, 2/3 d'apprentissage, 1/3 de test. Ceci peut être répété plusieurs fois afin d'avoir une idée, intra-jeu de données, de la robustesse de la prédiction par validation croisée.



Figure 41 : Photo de six porcelets LW nouveau-nés.
Ces six porcelets sont représentatifs de l'hétérogénéité de portée qui peut exister dans les lignées LW.

Jusqu'à présent, j'avais pu participer à trois essais de sélection de possibles prédicteurs :

- La recherche de gènes discriminant le caractère tendreté, chapitre 3c (non publiée).
- L'utilisation de la RMN pour identifier des *buckets* discriminants pour des phénotypes de production et de qualité des produits, chapitre 2, Figure 14 (Rohart *et al.* 2012).
- L'utilisation de la RMN pour identifier des *buckets* pour la sensibilité au stress thermique (Dou *et al.* 2017).

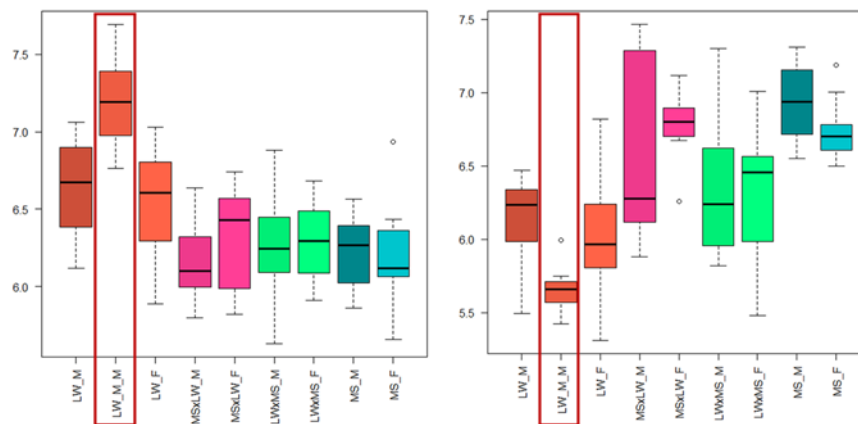


Figure 42 : Exemple de deux gènes surexprimé ou sous-exprimé que chez les mâles LW morts (LW_M_M) dans les 3 jours après la naissance.

L'expression des gènes discriminants chez les nouveau-nés qui vont mourir est encadrée en rouge.

Lors du projet PORCINET, des échantillons de sang total ont été prélevés à la naissance par Laurianne Canario et l'équipe du Magneraud où sont produits nos animaux expérimentaux. Le sang a été prélevé au cordon dans les minutes qui suivent la naissance, les animaux ont été phénotypés comme décrit dans la partie 1d des Perspectives. Une analyse du transcriptome du sang total a été réalisée pour 96 animaux des 4 génotypes, des 2 sexes et 8 LW mâles morts dans les trois 1ers jours après la naissance. Il n'y avait pas eu d'analyse du métabolome. A partir des données du transcriptome sur le sang naissance, 20 gènes avaient été identifiés par une simple analyse discriminante (Golub *et al.* 1999) ; l'exemple de deux gènes est présenté dans la figure 42 pour la mortalité avant trois jours chez les mâles LW (population plus à risque). Le premier gène surexprimé chez les nouveau-nés qui vont mourir est un gène impliqué dans l'homéostasie du fer décrit comme essentiel en période périnatale. Ce gène est déjà identifié comme biomarqueur dans des cas de sepsis néonatal précoce. Le deuxième gène, sous-exprimé chez ceux qui vont mourir, n'a pas de fonction connue en lien avec la survie néonatale. Il aurait une implication dans la régulation post-traductionnelle des protéines.

Pour les phénotypes quantitatifs, Laurianne Canario a réalisé une sélection de variables par sPLS qui n'a donné de résultats intéressants qu'après avoir fait trois groupes de phénotypes : poids, maturité,

comportement/mobilité. La Figure 43 présente les gènes sélectionnés (numéros) covariant avec les phénotypes en lien avec la mobilité. Vingt gènes étaient sélectionnés sur trois composantes, donc 60 gènes par groupe de phénotypes. Une liste de 180 gènes a été déterminée pour les trois groupes de phénotypes. Alors qu'on cherche une signature moléculaire dans le sang total, nous recherchons d'abord des gènes dont les modalités d'expression reflètent le mieux la variabilité phénotypique. Et comme l'exemple pour la tendreté de la viande, ces gènes peuvent être de fonction inconnue ou de fonction dont le lien avec le phénotype n'est pas évident. De plus, nous ne sommes pas dans le tissu le plus approprié pour la mobilité *a priori*. Il a donc été assez surprenant de trouver des gènes avec une certaine homogénéité fonctionnelle. Quelques exemples sont donnés sur la figure 43 où les gènes sélectionnés semblent préférentiellement correspondre à des fonctions musculaires, osseuses et de développement.

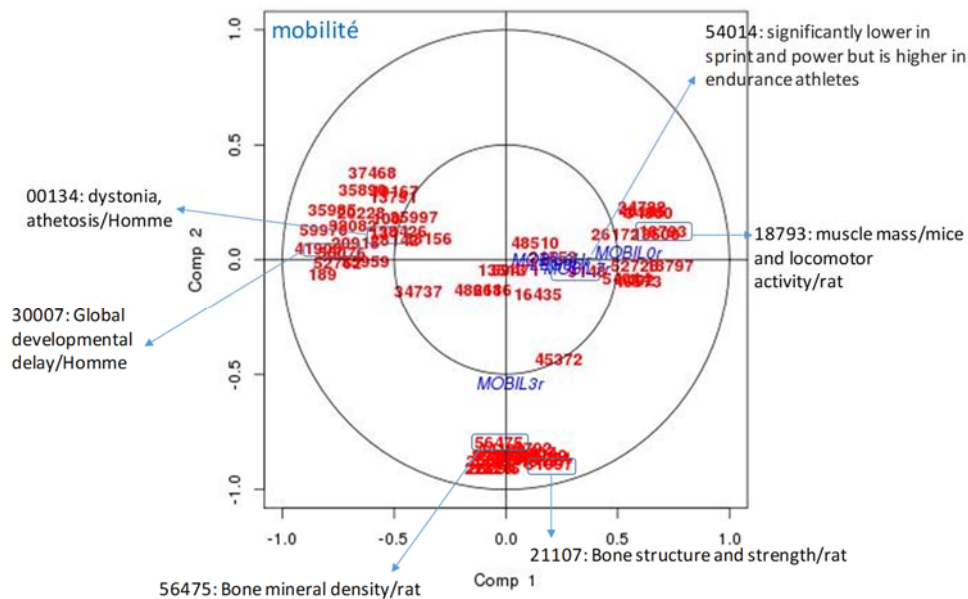


Figure 43 : Présentation des résultats de la sPLS pour le groupe de phénotypes relatif à la mobilité entre la naissance et 7 jours après.
Les gènes sont représentés par des numéros en rouge et les phénotypes en bleu.

Malheureusement et malgré ces résultats prometteurs, utiliser les expressions de gènes (dans notre cas) comme prédicteur nécessite de changer de méthode d'analyse de l'expression et de passer à de la PCR quantitative (qPCR) dite en temps réel (real-time qPCR). Nous utilisons la technologie Fluidigm BioMark HD disponible sur la plateforme GeT-Plage qui permet d'analyser l'expression de 96 gènes sur 96 échantillons (9216 points PCR). Il y a une contrainte forte, les amplifications doivent toutes se faire dans exactement les mêmes conditions d'amplification. Il y a aussi un problème spécifique à l'espèce porcine, qui est la qualité de l'annotation du génome du porc. Depuis 2018, nous disposons enfin d'une meilleure annotation néanmoins pas parfaite si bien que les essais peuvent s'avérer souvent infructueux. Un autre problème est que le transcriptome du sang total est celui qui nous donne le plus de difficultés d'analyse (signal global plus faible). Je passe les détails, deux ans de mises au point (test des amorces en PCR classique, test des amorces pour déterminer l'efficacité d'amplification, test des amorces sur le jeu de données PORCINET initial) pour avoir une liste de 66 gènes répondant à tous les critères techniques et de corrélation avec les phénotypes à partir d'une liste de 180 gènes testés. A partir des résultats des analyses d'expression de ces 66 gènes sur les données naissance initiales, il faut refaire une série d'analyses statistiques pour sélectionner les 45 gènes à utiliser en double (plus des gènes de contrôle). Les tests statistiques ont été réalisés sur tous les génotypes ou bien que sur les LW ; les échantillons indépendants ayant été prélevés que sur des animaux LW. Il s'agit d'optimiser le jeu final de gènes pour être corrélés aux phénotypes bruts ou corrigés. Sur les 66 gènes, des mesures de corrélation significative avec au moins un phénotype, des modèles linéaires simples, PLS, sPLS, sPLS-DA pour sélectionner des gènes qui peuvent ensuite être testés en régression logistique pour les phénotypes qualitatifs, et des arbres de décision. Ainsi un sous-jeu de 45 gènes a été défini pour être

testées sur les dispositifs expérimentaux indépendants.

L'autre approche en cours d'évaluation est d'utiliser une analyse métabolomique par RMN sur des échantillons de plasma ou de sérum prélevés à la naissance. Aucune analyse n'avait été réalisée dans le cadre de PORCINET sur des prélèvements naissance (seulement sur les fœtus). Nous avons toutefois des échantillons disponibles. La stratégie est donc différente ici ; pas de variables présélectionnées à tester sur des échantillons indépendants. Il s'agira donc d'avoir plusieurs dispositifs indépendants et de tester la prédiction des phénotypes. Les données générées seront testées soit sous forme de *buckets* dont les propriétés discriminantes ont déjà été démontrées (800 variables testées), soit sous forme de métabolites (190 variables testées) après transformation des spectres bruts en métabolites identifiés et quantifiés avec la méthode ASICS. La Figure 44 peut donner un aperçu de ce qui peut être obtenu. Il s'agit d'un réseau IPA obtenu avec les premiers résultats ASICS sur les données plasmatiques en fin de gestation. Ce réseau fait le lien avec des fonctions biologiques relatives à la question de la survie néonatale : stockage du glycogène, croissance, survie et comportement. Certaines de ces fonctions devraient être aussi représentées à la naissance. **Il s'agira donc d'établir un statut métabolique à la naissance et sa variabilité en fonction des différents dispositifs testés (robustesse du modèle) et de la variabilité phénotypique.**

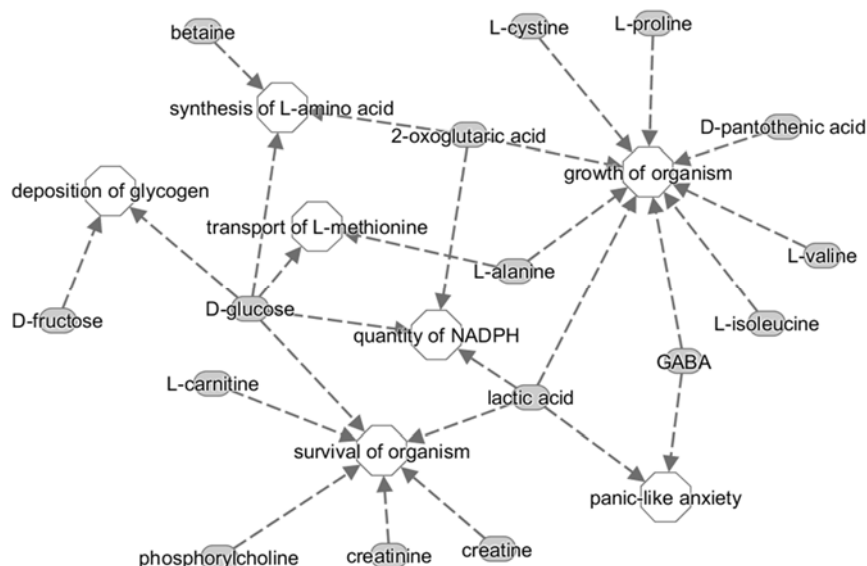


Figure 44 : Réseau IPA obtenu avec les données des métabolites plasmatiques en fin de gestation. Ce résultat a été obtenu avec les premières données RMN traduites en données quantitatives avec la méthode ASICS.

Au final, et sans prendre en compte le dispositif PORCINET à l'origine des données du transcriptome du sang total, nous disposons de 640 échantillons (plasma ou sérum, vraie naissance ou +10/20h) pour les études d'expression et de 615 échantillons pour les études de RMN. Ces prélèvements viennent de 170 nouveau-nés des deux sexes du dispositif PORCINET (LW, MSLW, LWMS, MS), de 190 nouveau-nés du dispositif SusOStress (lignées LW divergentes G3 pour la réponse de l'axe corticotrope à une injection d'ACTH), de 120 nouveau-nés du dispositif CMJR (lignées LW divergentes G10 pour l'efficacité alimentaire), et 110 nouveau-nés LW venant d'une maternité privée d'une OSP qui fait partie du regroupement Alliance R&D partenaire du projet SuBPig. Ce dernier lot correspond à des prélèvements « terrains » pour éprouver notre procédure, du prélèvement au résultat.

Nous avons donc un dispositif de 590 nouveau-nés avec parfois des prélèvements effectués dans les minutes qui suivent la naissance avec des mesures phénotypiques : poids, température rectale, longueur cou-queue, circonférence thorax, largeur épaule-épaule, mobilité dans la caisse, vocalisation, respiration, la mortalité avant sevrage. Certains porcelets ont également été prélevés au moment de la pesée officielle. Certains porcelets n'ont été prélevés qu'au moment de la pesée officielle (animaux

OSP).

Les approches statistiques sont toujours les mêmes plus celles que nous voudrions évaluer en fonction des résultats. Il s'agit toujours d'avoir des approches exploratoires avec des ACP, des PLS et des analyses de corrélation. Puis il s'agira d'évaluer des modèles type glm, des approches de types Forêts Aléatoires qui présentent l'avantage de pouvoir être utilisées en régression ou en analyse discriminante, des sPLS et sPLS-DA. Ce seront les données qui décideront de la stratégie pour définir le ou les modèles pour chaque phénotype et aussi selon le type de variable (gènes ou métabolites/*buckets*). Certains phénotypes présenteront plus de difficultés que d'autres. Si on prend l'exemple de la mortalité, les porcelets ne meurent pas tous pour les mêmes raisons et certains qui présentent des caractéristiques de morbidité survivent ; les données sont donc très bruitées. Sans compter que, heureusement, qu'une minorité de porcelets meurent (14% en moyenne) donc le dispositif pour les statistiques est très déséquilibré.

Nous menons donc parallèlement deux stratégies. Une stratégie où nous disposons de variables (des gènes) présélectionnées pour être discriminantes pour au moins un phénotype avec une seconde liste de gènes de secours (ils font partie des 66 gènes qui ont passé tous les tests). Et une stratégie ouverte avec les profils RMN où donc les variables ne sont pas présélectionnées si ce n'est le fait de tester une librairie de 190 métabolites avec ASICS. Les profils peuvent être traités avec la transformation en *buckets* ce qui étend le nombre possible de variables. Les deux approches, expression de gènes et profils métaboliques, seront comparées et pourront s'avérer complémentaires. Cette complémentarité pourrait permettre de surmonter le possible impact de la variabilité génétique qui peut se présenter dans les dispositifs présents et à venir. La Figure 45 présente ce qui est envisagé comme futur projet en génétique quantitative où les phénotypes moléculaires seraient mesurés à la naissance pour évaluer un ensemble de phénotypes qui seraient impossible à mesurer à la naissance et dans les jours qui suivent sur de grands dispositifs expérimentaux. L'idée est donc qu'à partir d'un prélèvement sanguin, plusieurs phénotypes sont prédits.

Il sera également intéressant de mettre en relation les phénotypes moléculaires mesurés précocement avec d'autres phénotypes mesurés en croissance ou à l'abattoir dans une seconde perspective de prédiction précoce de phénotypes tardifs.

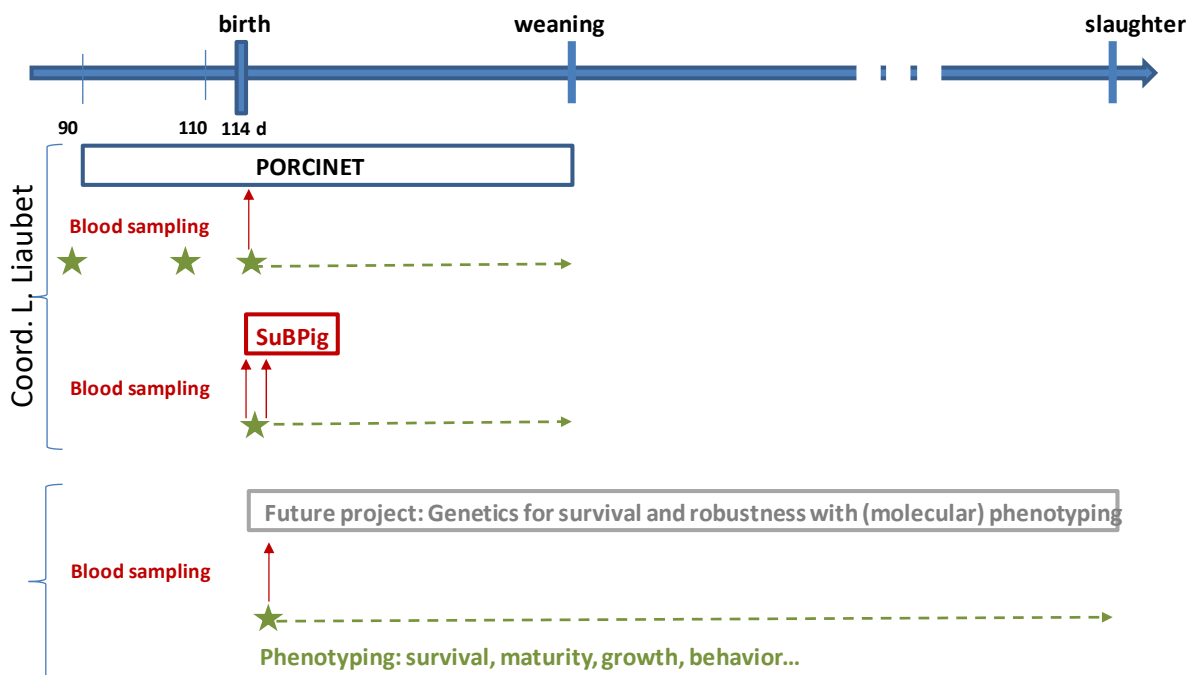


Figure 45 : Mise en perspective des projets PORCINET et SuBPig.

Une perspective possible est l'utilisation des biomarqueurs/prédicteurs comme phénotypes moléculaires dans un projet de génétique quantitative.

Au final, il s'agira de fournir des indicateurs moléculaires aux généticiens quantitatifs dans le cadre du développement d'un projet de génétique ; je ne serai pas coordinatrice d'un tel projet car hors de mes compétences.

A un niveau intermédiaire, il est par contre essentiel de continuer à collecter des échantillons afin de continuer à tester voire perfectionner nos modèles de prédiction, de les éprouver dans des contextes génétiques différents. En dehors de la génétique, il me semble qu'il serait intéressant de tester nos prédicteurs dans des contextes expérimentaux où on évalue des stratégies alimentaires/nutritionnelles sur la truie en gestation. Dit autrement, des stratégies nutritionnelles pourraient aussi être développées selon des paramètres déterminés à partir de nos données en gestation : nos données sur les fœtus et en particulier ce qui concerne le retard de développement de certains porcelets LW ainsi que nos données sur les interactions fœto-maternelles déjà acquises ou à acquérir. Pour cela, des collaborations sont à développer avec des physiologistes et des nutritionnistes.

A un niveau encore plus intégratif, l'ensemble de ces données et de celles dont je n'ai pas parlé des projets où j'apporte mon expertise sur l'analyse des données -omiques, servent à préciser les bases moléculaires de la notion de robustesse. L'ensemble de ces données peuvent servir à comprendre les interactions entre grandes fonctions et comment la sélection génétique a et pourra modifier les caractères. L'objectif est maintenant d'élever des animaux plus robustes dans une perspective d'agriculture durable, éthiquement plus acceptable.

4. Intégration génétique - génomique

Ne serait-ce qu'avec les données présentées dans ce document, nous disposons désormais de bases de données de génomique fonctionnelle (essentiellement des données d'expression) provenant de trois gros projets de recherche focalisés sur le muscle : fin de gestation, eQTL et diversité génétique. Par ailleurs, pour des projets auxquels je collabore, le transcriptome du sang total a été obtenu pour des nombres d'animaux importants dans des dispositifs QTL où donc des recherches d'eQTL pourront être menées. Il s'agit de projets de génétique sur les réponses de stress ou l'adaptation à la chaleur. Et dans tous ces projets, des données métabolomiques toutes obtenues avec des approches RMN sont également disponibles.

Cette ressource constituée d'informations fonctionnelles de la régulation du génome porcin à grande échelle pourrait être exploitée au-delà des questions propres aux projets d'où viennent les données. Mon outil d'intégration privilégié pour cet objectif est l'approche réseau.

Dans le cadre d'un autre projet de génétique chez le canard, j'ai réalisé des réseaux de co-occurrence simple sur des données en post-analyse génétique, ici sur des données pQTL où avait été calculé des QTL pour des caractères d'intérêt et des pQTL pour des abondances de protéines mesurées par une approche de protéomique (François et al. 2017). Dans cette étude avait également été testé le possible effet pléiotrope d'un QTL sur plusieurs phénotypes (qualité des produits, carcasse, cortisol plasmatique, abondance de protéines...) avec le test CLIP développé dans le laboratoire (David et al. 2013). Le graphe inféré reprend donc les résultats de liaison génétique où il est possible de visualiser si l'hypothèse de pléiotropie n'est pas rejetée. Le graphe obtenu ne fait donc que transformer de grands tableaux en une représentation simple et visuelle des résultats (Figure 6 du papier) où, en plus de la pléiotropie, se visualise facilement comment deux régions chromosomiques impactent à la fois plusieurs caractères. De nouveaux niveaux d'interaction sont suggérés (épistasie ?). L'extrait ci-dessous (Figure 46) illustre un résultat qui met dans un même sous-réseau, la corticostérone basale et le poids à 12 jours, huit protéines, dont une (MDH1) est génétiquement lié aux deux phénotypes. Seul le QTL sur le chromosome 9 satisfait le test de pléiotropie pour la régulation de l'abondance d'une protéine (FABP7) et d'un phénotype (BW12).

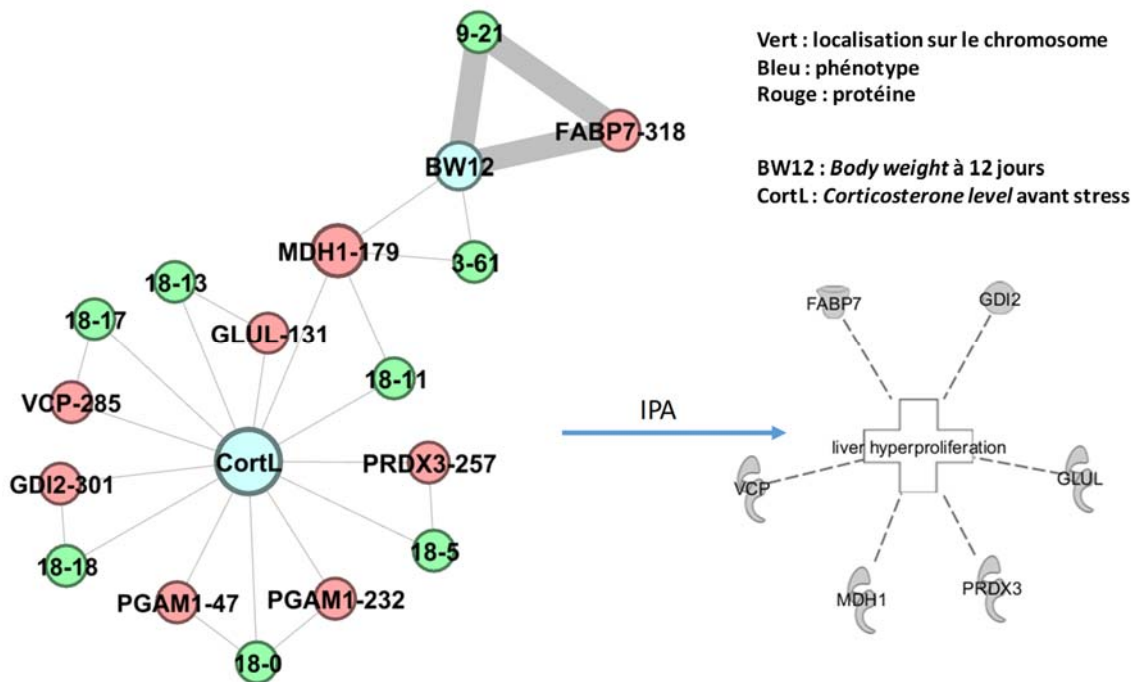


Figure 46 : Graphe obtenu par inférence simple après détection de QTL et de pQTL et test de la pléiotropie. Ce graphe est juste le résultat de la transformation post analyse QTL d'un tableau en graphe pour faciliter l'interprétation. Ici les données protéiques du graphe ont ensuite été soumises à IPA pour aide à l'interprétation.

Pour ce même projet, nous disposons également de l'information au niveau expression de gènes et RMN. Pour l'inférence de graphes, je ne retiens plus que les interactions qui satisfont le test de pléiotropie. Le graphe est quand même énorme (non montré). Nous testons aussi des approches PLS avec transformation en réseau bipartite. Ou bien encore nous pouvons aussi envisager une approche similaire à celle de la figure 8 de ce document. Des auteurs suggèrent que l'épistasie et la pléiotropie seraient des propriétés inhérentes aux réseaux biomoléculaires plutôt que des occurrences isolées, ce qui justifierait la nécessité d'une compréhension de la génétique par une approche systèmes (Braun 2014).

Nous pouvons aussi comparer des graphes de co-expression (corrélation partielle) à partir d'un même lot de gènes d'intérêt obtenu dans des dispositifs expérimentaux différents selon l'exemple donné dans la figure 16 de ce document. Nous pouvons aussi intégrer de l'information phénotypique comme dans la figure 8 ou les figures 27 et 28 de ce document...

Ces résultats peuvent ensuite intégrer l'information génétique disponible, soit en utilisant les associations génétiques pour générer un graphe, soit en recherchant de l'information génétique sur les gènes révélés importants dans les graphes précédents.

Une autre approche peut être de s'intéresser aux 80 gènes de la voie de signalisation des PI3K (Figure 15 où ne sont représentés que 40 gènes) au cœur de nos résultats sur la diversité génétique entre les lignées pures utilisées dans les croisements en production. Il s'agirait de faire le lien avec l'information génétique disponible pour ces gènes (co-localisation avec des QTL, régulés par un ou des eQTL...).

Et puis encore, dans une démarche inverse, des régions d'intérêt sont identifiées pour réguler des phénotypes, des traces de sélection définissent aussi des régions génomiques. Les gènes de ces régions peuvent être recherchés dans les différents réseaux disponibles pour évaluer leur pertinence par rapport à la question génétique et améliorer la sélection de gènes candidats.

Un projet de recherche s'intéressant justement à ce type d'approches (et d'autres) a été financé par le Métaprogramme Selgen et est coordonné par Juliette Riquet. Du point de vue du généticien, les études QTL, GWAS et traces de sélection donnent des régions génomiques d'intérêt et la première approche est de s'intéresser au gène le plus proche du marqueur génétique (SNP surtout). Finalement peu de mutations causales ont été ainsi caractérisées. C'est pourquoi l'idée de combiner de l'information

fonctionnelle de différentes origines (expression, *gene ontology*, HiC...) permettrait de gagner en puissance dans l'identification des gènes candidats et dans la priorisation de leur possible implication pour la question posée. Un exemple de ce type d'approche est décrit dans Tasan et al. (2015). Les auteurs proposent de construire des réseaux de co-fonctions pour identifier des gènes mutuellement liés sur plusieurs loci.

Les travaux de biologie intégrative pour la génétique demandent des interactions fortes entre disciplines différentes. Chacune de ces disciplines a une marche de progression différente avec ses propres exigences. Ce qui me donne le sentiment d'avancer plus lentement que la recherche sur projet ne le demande. Malgré tout j'ai le sentiment de participer à des projets ambitieux. Ma part est de démontrer comment les approches de génomique fonctionnelle peuvent permettre de faire avancer les projets de génétique.

Références

1. Battaglia F.C. & Meschia G. (1978) Principal substrates of fetal metabolism. *Physiol Rev* 58, 499-527.
2. Bidanel J.P., Milan D., Iannuccelli N., Amigues Y., Boscher M.Y., Bourgeois F., Caritez J.C., Gruand J., Le Roy P., Lagant H., Quintanilla R., Renard C., Gellin J., Ollivier L. & Chevalet C. (2001) Detection of quantitative trait loci for growth and fatness in pigs. *Genet Sel Evol* 33, 289-309.
3. Biensen N.J., Wilson M.E. & Ford S.P. (1999) The impacts of uterine environment and fetal genotype on conceptus size and placental vascularity during late gestation in pigs. *J Anim Sci* 77, 954-9.
4. Bischoff S.R., Tsai S., Hardison N., Motsinger-Reif A.A., Freking B.A., Nonneman D., Rohrer G. & Piedrahita J.A. (2009) Characterization of conserved and nonconserved imprinted genes in swine. *Biol Reprod* 81, 906-20.
5. Blaauw B., Schiaffino S. & Reggiani C. (2013) Mechanisms modulating skeletal muscle phenotype. *Compr Physiol* 3, 1645-87.
6. Blum Y., Le Mignon G., Causeur D., Filangi O., Desert C., Demeure O., Le Roy P. & Lagarrigue S. (2011) Complex trait subtypes identification using transcriptome profiling reveals an interaction between two QTL affecting adiposity in chicken. *BMC Genomics* 12, 567.
7. Bonnet A., Lagarrigue S., Liaubet L., Robert-Granié C., Sancristobal M. & Tosser-Klopp G. (2009) Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares. *BMC Proc* 3 Suppl 4, S11.
8. Bonnet A., Le Cao K.A., Sancristobal M., Benne F., Robert-Granié C., Law-So G., Fabre S., Besse P., De Billy E., Quesnel H., Hatey F. & Tosser-Klopp G. (2008) In vivo gene expression in granulosa cells during pig terminal follicular development. *Reproduction* 136, 211-24.
9. Braun R. (2014) Systems analysis of high-throughput data. *Adv Exp Med Biol* 844, 153-87.
10. Breiman L. (2001) Random Forests. *Machine Learning* 45, 5-32.
11. Brem R.B., Yvert G., Clinton R. & Kruglyak L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752-5.
12. Brown D.M., Williams H., Ryan K.J., Wilson T.L., Daniel Z.C., Mareko M.H., Emes R.D., Harris D.W., Jones S., Wattis J.A., Dryden I.L., Hodgman T.C., Brameld J.M. & Parr T. (2016) Mitochondrial phosphoenolpyruvate carboxykinase (PEPCK-M) and serine biosynthetic pathway genes are co-ordinately increased during anabolic agent-induced skeletal muscle growth. *Sci Rep* 6, 28693.
13. Byrne K., Vuocolo T., Gondro C., White J.D., Cockett N.E., Hadfield T., Bidwell C.A., Waddell J.N. & Tellam R.L. (2010) A gene network switch enhances the oxidative capacity of ovine skeletal muscle during late fetal development. *BMC Genomics* 11, 378.
14. Canario L., Cantoni E., Le Bihan E., Caritez J.C., Billon Y., Bidanel J.P. & Foulley J.L. (2006) Between-breed variability of stillbirth and its relationship with sow and piglet characteristics. *J Anim Sci* 84, 3185-96.
15. Canario L., Père M.-C., Quesnel H., Billon Y., Herbrard W., Riquet J., Mormède P. & Liaubet L. (2014) Influences génétiques - Large White et Meishan - sur la fin du développement de foetus purs et croisés de la même portée. pp. 25-30. *Journées Recherche Porcine*.
16. Canario L., Père M.C., Tribout T., Thomas F., David C., Gogué J., Herpin P., Bidanel J.P. & Le Dividich J. (2007) Estimation of genetic trends from 1977 to 1998 of body composition and physiological state of Large White pigs at birth. *Animal* 1, 1409-13.
17. Casarin A., Jimenez-Ortega J.C., Trevisson E., Pertegato V., Doimo M., Ferrero-Gomez M.L., Abbadì S., Artuch R., Quinzii C., Hirano M., Basso G., Ocana C.S., Navas P. & Salviati L. (2008) Functional characterization of human COQ4, a gene required for Coenzyme Q10 biosynthesis. *Biochem Biophys Res Commun* 372, 35-9.
18. Chen C., Guo Y., Yang G., Yang Z., Zhang Z., Yang B., Yan X., Perez-Enciso M., Ma J., Duan Y., Brenig B. & Huang L. (2009) A genome wide detection of quantitative trait loci on pig maternal infanticide behavior in a large scale White Duroc x Erhualian resource population. *Behav Genet* 39, 213-9.

19. Cherel P., Herault F., Vincent A., Le Roy P. & Damon M. (2012) Genetic variability of transcript abundance in pig skeletal muscle at slaughter: relationships with meat quality traits. *J Anim Sci* 90, 699-708.
20. Cherel P., Pires J., Glenisson J., Milan D., Iannuccelli N., Herault F., Damon M. & Le Roy P. (2011) Joint analysis of quantitative trait loci and major-effect causative mutations affecting meat quality and carcass composition traits in pigs. *BMC Genet* 12, 76.
21. Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibé B., Bouix J., Caiment F., Elsen J.M., Eychenne F., Larzul C., Laville E., Meish F., Milenkovic D., Tobin J., Charlier C. & Georges M. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* 38, 813-8.
22. Congras A., Yerle-Bouissou M., Pinton A., Vignoles F., Liaubet L., Ferchaud S. & Acloque H. (2014) Sperm DNA methylation analysis in swine reveals conserved and species-specific methylation patterns and highlights an altered methylation at the GNAS locus in infertile boars. *Biol Reprod* 91, 137.
23. David I., Elsen J.M. & Concordet D. (2013) CLIP Test: a new fast, simple and powerful method to distinguish between linked or pleiotropic quantitative trait loci in linkage disequilibria analysis. *Heredity (Edinb)* 110, 232-8.
24. de Koning D.J., Harlizius B., Rattink A.P., Groenen M.A., Brascamp E.W. & van Arendonk J.A. (2001) Detection and characterization of quantitative trait loci for meat quality traits in pigs. *J Anim Sci* 79, 2812-9.
25. de Koning D.J., Janss L.L., Rattink A.P., van Oers P.A., de Vries B.J., Groenen M.A., van der Poel J.J., de Groot P.N., Brascamp E.W. & van Arendonk J.A. (1999) Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). *Genetics* 152, 1679-90.
26. de Vienne D., Leonardi A. & Damerval C. (1988) Genetic aspects of variation of protein amounts in maize and pea. *Electrophoresis* 9, 742-50.
27. Demeure O., Liaubet L., Riquet J. & Milan D. (2004) Determination of PRKAG1 coding sequence and mapping of PRKAG1 and PRKAG2 relatively to porcine back fat thickness QTL. *Anim Genet* 35, 123-5.
28. Deslandes L., Pileur F., Liaubet L., Camut S., Can C., Williams K., Holub E., Beynon J., Arlat M. & Marco Y. (1998) Genetic characterization of RRS1, a recessive locus in *Arabidopsis thaliana* that confers resistance to the bacterial soilborne pathogen *Ralstonia solanacearum*. *Mol Plant Microbe Interact* 11, 659-67.
29. Dou S., Villa-Vialaneix N., Liaubet L., Billon Y., Giorgi M., Gilbert H., Gourdine J.L., Riquet J. & Renaudeau D. (2017) 1HNMR-Based metabolomic profiling method to develop plasma biomarkers for sensitivity to chronic heat stress in growing pigs. *PLoS One* 12, e0188469.
30. Drag M., Hansen M.B. & Kadarmideen H.N. (2018) Systems genomics study reveals expression quantitative trait loci, regulator genes and pathways associated with boar taint in pigs. *PLoS One* 13, e0192673.
31. Egerman M.A. & Glass D.J. (2014) Signaling pathways controlling skeletal muscle mass. *Crit Rev Biochem Mol Biol* 49, 59-68.
32. Ferre P.J., Liaubet L., Concordet D., SanCristobal M., Uro-Coste E., Tosser-Klopp G., Bonnet A., Toutain P.L., Hately F. & Lefebvre H.P. (2007) Longitudinal analysis of gene expression in porcine skeletal muscle after post-injection local injury. *Pharm Res* 24, 1480-9.
33. Foxcroft G.R., Dixon W.T., Novak S., Putnam C.T., Town S.C. & Vinsky M.D. (2006) The biological basis for prenatal programming of postnatal performance in pigs. *J Anim Sci* 84 Suppl, E105-12.
34. Frantz C., Stewart K.M. & Weaver V.M. (2010) The extracellular matrix at a glance. *J Cell Sci* 123, 4195-200.
35. François Y., Vignal A., Molette C., Marty-Gasset N., Davail S., Liaubet L. & Marie-Etancelin C. (2017) Deciphering mechanisms underlying the genetic variation of general production and liver quality traits in the overfed mule duck by pQTL analyses. *Genet Sel Evol* 49, 38.
36. Gaibelet G., Meilhoc E., Riond J., Saves I., Exner T., Liaubet L., Nürnberg B., Masson J.M. & Emorine L.J. (1999) Nonselective coupling of the human mu-opioid receptor to multiple inhibitory G-protein isoforms. *Eur J Biochem* 261, 517-23.

37. Gilad Y., Oshlack A. & Rifkin S.A. (2006) Natural selection on gene expression. *Trends Genet* 22, 456-61.
38. Gilad Y., Rifkin S.A. & Pritchard J.K. (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet* 24, 408-15.
39. Glass D.J. (2003) Molecular mechanisms modulating muscle mass. *Trends Mol Med* 9, 344-50.
40. Godiard L., Sauviac L., Dalbin N., Liaubet L., Callard D., Czernic P. & Marco Y. (1998) CYP76C2, an *Arabidopsis thaliana* cytochrome P450 gene expressed during hypersensitive and developmental cell death. *FEBS Lett* 438, 245-9.
41. Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Coller H., Loh M.L., Downing J.R., Caligiuri M.A., Bloomfield C.D. & Lander E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-7.
42. Gondret F., Guével B., Père M.C., Quesnel H., Billon Y., Com E., Canario L., Louveau I. & Liaubet L. (2018) Proteomic analysis of adipose tissue during the last weeks of gestation in pure and crossbred Large White or Meishan fetuses gestated by sows of either breed. *J Anim Sci Biotechnol* 9, 28.
43. Gonzalez-Prendes R., Quintanilla R. & Amills M. (2017) Investigating the genetic regulation of the expression of 63 lipid metabolism genes in the pig skeletal muscle. *Anim Genet* 48, 606-10.
44. Gonzalez-Prendes R., Quintanilla R., Marmol-Sanchez E., Pena R.N., Ballester M., Cardoso T.F., Manunza A., Casellas J., Canovas A., Diaz I., Noguera J.L., Castello A., Mercade A. & Amills M. (2019) Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine gluteus medius and longissimus dorsi muscles. *BMC Genomics* 20, 170.
45. Gorni C., Garino C., Iacuaniello S., Castiglioni B., Stella A., Restelli G.L., Pagnacco G. & Mariani P. (2011) Transcriptome analysis to identify differential gene expression affecting meat quality in heavy Italian pigs. *Anim Genet* 42, 161-71.
46. Gorni C., Iacuaniello S., Castiglioni B., Pagnacco G. & Mariani P. (2008) Pig KALRN, MYH1, MLC2V, SNX13, AK1, and PPIA loci RH mapping and chromosome position refining. *Genet Mol Res* 7, 982-5.
47. Guerrero-Bosagna C., Morisson M., Liaubet L., Rodenburg T.B., de Haas E.N., Košťál L. & Pitel F. (2018) Transgenerational epigenetic inheritance in birds. *Environ Epigenet* 4, dvy008.
48. Guo B., Greenwood P.L., Cafe L.M., Zhou G., Zhang W. & Dalrymple B.P. (2015) Transcriptome analysis of cattle muscle identifies potential markers for skeletal muscle growth rate and major cell types. *BMC Genomics* 16, 177.
49. Guo L., Qiao M., Wang C., Zheng R., Xiong Y.Z. & Deng C.Y. (2012) Imprinting analysis of porcine MAGEL2 gene in two fetal stages and association analysis with carcass traits. *Mol Biol Rep* 39, 147-55.
50. Hazard D., Liaubet L., Sancristobal M. & Mormède P. (2008) Gene array and real time PCR analysis of the adrenal sensitivity to adrenocorticotrophic hormone in pig. *BMC Genomics* 9, 101.
51. Hedegaard J., Arce C., Bicciato S., Bonnet A., Buitenhuis B., Collado-Romero M., Conley L.N., SanCristobal M., Ferrari F., Garrido J.J., Groenen M.A., Hornshøj H., Hulsegge I., Jiang L., Jiménez-Marín Á., Kommadath A., Lagarrigue S., Leunissen J.A., Liaubet L., Neerinx P.B., Nie H., Poel J.V., Prickett D., Ramirez-Boo M., Rebel J.M., Robert-Granié C., Skarman A., Smits M.A., Sørensen P., Tosser-Klopp G. & Watson M. (2009) Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. *BMC Proc* 3 Suppl 4, S5.
52. Heidt H., Cinar M.U., Uddin M.J., Looft C., Jungst H., Tesfaye D., Becker A., Zimmer A., Ponsuksili S., Wimmers K., Tholen E., Schellander K. & Grosse-Brinkhaus C. (2013) A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population. *Mamm Genome* 24, 416-26.
53. Herpin P., Lossec G., Schmidt I., Cohen-Adad F., Duchamp C., Lefaucheur L., Goglia F. & Lanni A. (2002) Effect of age and cold exposure on morphofunctional characteristics of skeletal muscle in neonatal pigs. *Pflugers Arch* 444, 610-8.
54. Howland M.A., Sandman C.A. & Glynn L.M. (2017) Developmental origins of the human hypothalamic-pituitary-adrenal axis. *Expert Rev Endocrinol Metab* 12, 321-39.
55. Hu Z.L., Dracheva S., Jang W., Maglott D., Bastiaansen J., Rothschild M.F. & Reecy J.M. (2005) A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm Genome* 16, 792-800.

56. Hube F. & Francastel C. (2015) Mammalian introns: when the junk generates molecular diversity. *Int J Mol Sci* 16, 4429-52.
57. Huber K. (2018) Invited review: resource allocation mismatch as pathway to disproportionate growth in farm animals - prerequisite for a disturbed health. *Animal* 12, 528-36.
58. Hudson N.J., Lyons R.E., Reverter A., Greenwood P.L. & Dalrymple B.P. (2013) Inferring the in vivo cellular program of developing bovine skeletal muscle from expression data. *Gene Expr Patterns* 13, 109-25.
59. Jansen R.C. & Nap J.P. (2001) Genetical genomics: the added value from segregation. *Trends Genet* 17, 388-91.
60. Jeon J.T., Carlborg O., Tornsten A., Giuffra E., Amarger V., Chardon P., Andersson-Eklund L., Andersson K., Hansson I., Lundstrom K. & Andersson L. (1999) A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nat Genet* 21, 157-8.
61. Jiang C., Yang Y., Huang C. & Whitelaw B. (2014) Promoter characterization and functional association with placenta of porcine MAGEL2. *Gene* 547, 63-9.
62. Jorquera R., Ortiz R., Ossandon F., Cardenas J.P., Sepulveda R., Gonzalez C. & Holmes D.S. (2016) SinEx DB: a database for single exon coding sequences in mammalian genomes. *Database (Oxford)* 2016.
63. Jouffe V., Rowe S., Liaubet L., Buitenhuis B., Hornshøj H., Sancristobal M., Mormède P. & de Koning D.J. (2009) Using microarrays to identify positional candidate genes for QTL: the case study of ACTH response in pigs. *BMC Proc* 3 Suppl 4, S14.
64. Kadarmideen H.N., von Rohr P. & Janss L.L. (2006) From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mamm Genome* 17, 548-64.
65. Kambadur R., Sharma M., Smith T.P. & Bass J.J. (1997) Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Res* 7, 910-6.
66. Lacomme M., Liaubet L., Pituello F. & Bel-Vialar S. (2012) NEUROG2 drives cell cycle exit of neuronal precursors by specifically repressing a subset of cyclins acting at the G1 and S phases of the cell cycle. *Mol Cell Biol* 32, 2596-607.
67. Lahbib-Mansais Y., Barasc H., Marti-Marimon M., Mompert F., Iannuccelli E., Robelin D., Riquet J. & Yerle-Bouissou M. (2016) Expressed alleles of imprinted IGF2, DLK1 and MEG3 colocalize in 3D-preserved nuclei of porcine fetal cells. *BMC Cell Biol* 17, 35.
68. Laporte J., Liaubet L., Blondeau F., Tronchère H., Mandel J.L. & Payrastre B. (2002) Functional redundancy in the myotubularin family. *Biochem Biophys Res Commun* 291, 305-12.
69. Laville E., Sayd T., Terlouw C., Blinet S., Pinguet J., Fillaut M., Glenisson J. & Cherel P. (2009) Differences in pig muscle proteome according to HAL genotype: implications for meat quality defects. *J Agric Food Chem* 57, 4913-23.
70. Laville E., Sayd T., Terlouw C., Chambon C., Damon M., Larzul C., Leroy P., Glenisson J. & Cherel P. (2007) Comparison of sarcoplasmic proteomes between two groups of pig muscles selected for shear force of cooked meat. *J Agric Food Chem* 55, 5834-41.
71. Le Cao K.A., Boitard S. & Besse P. (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 12, 253.
72. Le Mignon G., Blum Y., Demeure O., Diot C., Le Bihan-Duval E., Le Roy P. & Lagarrigue S. (2010) Apports de la génomique fonctionnelle à la cartographie fine de QTL. *Productions Animales* 3, 343-58.
73. Leenhouwers J.I., Knol E.F., de Groot P.N., Vos H. & van der Lende T. (2002) Fetal development in the pig in relation to genetic merit for piglet survival. *J Anim Sci* 80, 1759-70.
74. Lefort G., Liaubet L., Canlet C., Tardivel P., Pere M.C., Quesnel H., Paris A., Iannuccelli N., Vialaneix N. & Servien R. (2019) ASICS: an R package for a whole analysis workflow of 1D 1H NMR spectra. *Bioinformatics*.
75. Lehnert S.A., Reverter A., Byrne K.A., Wang Y., Natrass G.S., Hudson N.J. & Greenwood P.L. (2007) Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Dev Biol* 7, 95.
76. Li H.D., Lund M.S., Christensen O.F., Gregersen V.R., Henckel P. & Bendixen C. (2010) Quantitative trait loci analysis of swine meat quality traits. *J Anim Sci* 88, 2904-12.
77. Liaubet L., Bertrand N., Medevielle F. & Pituello F. (2000) Identification by differential display

- of a chicken tolloid-related metalloprotease specifically expressed in the caudal notochord. *Mech Dev* 96, 101-5.
78. Liaubet L., Lobjois V., Faraut T., Tircazes A., Benne F., Iannuccelli N., Pires J., Glenisson J., Robic A., Le Roy P., Sancristobal M. & Cherel P. (2011) 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* 12, 548.
 79. Listrat A., Leuret B., Louveau I., Astruc T., Bonnet M., Lefaucheur L., Picard B. & Bugeon J. (2016) How Muscle Structure and Composition Influence Meat and Flesh Quality. *ScientificWorldJournal* 2016, 3182746.
 80. Liu T.Y., Chen Y.C., Jong Y.J., Tsai H.J., Lee C.C., Chang Y.S., Chang J.G. & Chang Y.F. (2017) Muscle developmental defects in heterogeneous nuclear Ribonucleoprotein A1 knockout mice. *Open Biol* 7.
 81. Lobjois V., Liaubet L., SanCristobal M., Glénisson J., Fève K., Rallières J., Le Roy P., Milan D., Cherel P. & Hately F. (2008) A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. *Anim Genet* 39, 147-62.
 82. Lunde A., Melve K.K., Gjessing H.K., Skjaerven R. & Irgens L.M. (2007) Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. *Am J Epidemiol* 165, 734-41.
 83. Ma J., Yang J., Zhou L., Ren J., Liu X., Zhang H., Yang B., Zhang Z., Ma H., Xie X., Xing Y., Guo Y. & Huang L. (2014) A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLoS Genet* 10, e1004710.
 84. Malek M., Dekkers J.C., Lee H.K., Baas T.J., Prusa K., Huff-Lonergan E. & Rothschild M.F. (2001) A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. II. Meat and muscle composition. *Mamm Genome* 12, 637-45.
 85. Maltin C., Balcerzak D., Tilley R. & Delday M. (2003) Determinants of meat quality: tenderness. *Proc Nutr Soc* 62, 337-47.
 86. Maroilley T., Lemonnier G., Lecardonnel J., Esquerré D., Ramayo-Caldas Y., Mercat M.J., Rogel-Gaillard C. & Estellé J. (2017) Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* 18, 967.
 87. Marti-Marimon M., Vialaneix N., Voillet V., Yerle-Bouissou M., Lahbib-Mansais Y. & Liaubet L. (2018) A new approach of gene co-expression network inference reveals significant biological processes involved in porcine muscle development in late gestation. *Sci Rep* 8, 10150.
 88. Mellor D.J. & Cockburn F. (1986) A comparison of energy metabolism in the new-born infant, piglet and lamb. *Q J Exp Physiol* 71, 361-79.
 89. Mesa H., Cammack K.M., Safranski T.J., Green J.A. & Lamberson W.R. (2012) Selection for placental efficiency in swine: conceptus development. *J Anim Sci* 90, 4217-22.
 90. Mohr S. & Liew C.C. (2007) The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol Med* 13, 422-32.
 91. Montastier E., Villa-Vialaneix N., Caspar-Bauguil S., Hlavaty P., Tvřizicka E., Gonzalez I., Saris W.H., Langin D., Kunesova M. & Viguierie N. (2015) System model network for adipose tissue signatures related to weight changes in response to calorie restriction and subsequent weight maintenance. *PLoS Comput Biol* 11, e1004047.
 92. Mormede P., Foury A., Terenina E. & Knap P.W. (2011) Breeding for robustness: the role of cortisol. *Animal* 5, 651-7.
 93. Morzel M., Chambon C., Hamelin M., Sante-Lhoutellier V., Sayd T. & Monin G. (2004) Proteome changes during pork meat ageing following use of two different pre-slaughter handling procedures. *Meat Science* 67, 689-96.
 94. Mota-Rojas D., Orozco-Gregorio H., Villanueva-Garcia D., Bonilla-Jaime H., Suarez-Bonilla X., Hernandez-Gonzalez R., Roldan-Santiago P. & Trujillo-Ortega M.E. (2011) Foetal and neonatal energy metabolism in pigs and humans: a review. pp. 215-25. *Veterinarni Medicina*.
 95. Nezer C., Moreau L., Brouwers B., Coppieters W., Detilleux J., Hanset R., Karim L., Kvasz A., Leroy P. & Georges M. (1999) An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat Genet* 21, 155-6.
 96. Oleksiak M.F., Churchill G.A. & Crawford D.L. (2002) Variation in gene expression within and

- among natural populations. *Nat Genet* 32, 261-6.
97. Orczyk-Pawilowicz M., Jawien E., Deja S., Hirnle L., Zabek A. & Mlynarz P. (2016) Metabolomics of Human Amniotic Fluid and Maternal Plasma during Normal Pregnancy. *PLoS One* 11, e0152740.
 98. Palmas F., Fattuoni C., Noto A., Barberini L., Dessi A. & Fanos V. (2016) The choice of amniotic fluid in metabolomics for the monitoring of fetus health. *Expert Rev Mol Diagn* 16, 473-86.
 99. Picard B., Berri C., Lefaucheur L., Molette C., Sayd T. & Terlouw C. (2010) Skeletal muscle proteomics in livestock production. *Brief Funct Genomics* 9, 259-78.
 100. Picard B., Lebret B., Cassar-Malek I., Liaubet L., Berri C., Le Bihan-Duval E., Hocquette J.F. & Renand G. (2015) Recent advances in omic technologies for meat quality management. *Meat Sci* 109, 18-26.
 101. Picard B., Lefaucheur L., Berri C. & Duclos M.J. (2002) Muscle fibre ontogenesis in farm animal species. *Reprod Nutr Dev* 42, 415-31.
 102. Rauw W.M. (2016) Editorial: Improving Animal Welfare through Genetic Selection. *Front Genet* 7, 69.
 103. Rauw W.M., Kanis E., Noordhuizen-Stassen E.N. & Grommers F.J. (1998) Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science* 56, 15-33.
 104. Rehfeldt C., Fiedler I., Dietl G. & Ender K. (2000) Myogenesis and postnatal skeletal muscle cell growth as influenced by selection. *Livestock Production Science* 66, 177-88.
 105. Ren H., Li L., Su H., Xu L., Wei C., Zhang L., Li H., Liu W. & Du L. (2011) Histological and transcriptome-wide level characteristics of fetal myofiber hyperplasia during the second half of gestation in Texel and Ujumqin sheep. *BMC Genomics* 12, 411.
 106. Reverter A. & Chan E.K. (2008) Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491-7.
 107. Robic A., Faraut T., Liaubet L., Riquet J., Milan D. & Lobjois V. (2008) Characterization of porcine ASB6 gene and transcripts-comparison of mammalian transcripts. *Anim Biotechnol* 19, 138-43.
 108. Rohart F., Paris A., Laurent B., Canlet C., Molina J., Mercat M.J., Tribout T., Muller N., Iannuccelli N., Villa-Vialaneix N., Liaubet L., Milan D. & San Cristobal M. (2012) Phenotypic prediction based on metabolomic data for growing pigs from three main European breeds. *J Anim Sci* 90, 4729-40.
 109. Rohrer G.A. & Keele J.W. (1998) Identification of quantitative trait loci affecting carcass composition in swine: II. Muscling and wholesale product yield traits. *J Anim Sci* 76, 2255-62.
 110. Rohrer G.A., Thallman R.M., Shackelford S., Wheeler T. & Koohmaraie M. (2006) A genome scan for loci affecting pork quality in a Duroc-Landrace F population. *Anim Genet* 37, 17-27.
 111. Rothschild M.F. (2003) From a sow's ear to a silk purse: real progress in porcine genomics. *Cytogenet Genome Res* 102, 95-9.
 112. Rothschild M.F. (2004) Porcine genomics delivers new tools and results: this little piggy did more than just go to market. *Genet Res* 83, 1-6.
 113. Rozario T. & DeSimone D.W. (2010) The extracellular matrix in development and morphogenesis: a dynamic view. *Dev Biol* 341, 126-40.
 114. Rubin C.J., Megens H.J., Martinez Barrio A., Maqbool K., Sayyab S., Schwochow D., Wang C., Carlborg Ö., Jern P., Jørgensen C.B., Archibald A.L., Fredholm M., Groenen M.A. & Andersson L. (2012) Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A* 109, 19529-36.
 115. Sanchez M.P., Riquet J., Iannuccelli N., Gogue J., Billon Y., Demeure O., Caritez J.C., Burgaud G., Feve K., Bonnet M., Pery C., Lagant H., Le Roy P., Bidanel J.P. & Milan D. (2006) Effects of quantitative trait loci on chromosomes 1, 2, 4, and 7 on growth, carcass, and meat quality traits in backcross Meishan x Large White pigs. *J Anim Sci* 84, 526-37.
 116. SanCristobal M., Rohart F., Lascor C., Bouffaud M., Trouilh L., Martin P.G., Lippi Y., Tribout T., Faraut T., Mercat M.J., Milan D. & Liaubet L. (2015) Exploring transcriptomic diversity in muscle revealed that cellular signaling pathways mainly differentiate five Western porcine breeds. *BMC Genomics* 16, 1055.

117. Sautron V., Terenina E., Gress L., Lippi Y., Billon Y., Larzul C., Liaubet L., Villa-Vialaneix N. & Mormède P. (2015) Time course of the response to ACTH in pig: biological and transcriptomic study. *BMC Genomics* 16, 961.
118. Sayd T., Morzel M., Chambon C., Franck M., Figwer P., Larzul C., Le Roy P., Monin G., Cherel P. & Laville E. (2006) Proteome analysis of the sarcoplasmic fraction of pig semimembranosus muscle: implications on meat color development. *J Agric Food Chem* 54, 2732-7.
119. 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., Linsley P.S., Mao M., Stoughton R.B. & Friend S.H. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297-302.
120. Shang Y., Zhang C., Wang S., Xiong F., Zhao C., Peng F., Feng S., Yu M., Li M. & Zhang Y. (2007) Activated beta-catenin induces myogenesis and inhibits adipogenesis in BM-derived mesenchymal stromal cells. *Cytotherapy* 9, 667-81.
121. Sivakumar A., de Las Heras J.I. & Schirmer E.C. (2019) Spatial Genome Organization: From Development to Disease. *Front Cell Dev Biol* 7, 18.
122. Sudre K., Leroux C., Piétu G., Cassar-Malek I., Petit E., Listrat A., Auffray C., Picard B., Martin P. & Hocquette J.F. (2003) Transcriptome analysis of two bovine muscles during ontogenesis. *J Biochem* 133, 745-56.
123. Szyda J., Grindflek E., Liu Z. & Lien S. (2003) Multivariate mixed inheritance models for QTL detection on porcine chromosome 6. *Genet Res* 81, 65-73.
124. Tacer K.F. & Potts P.R. (2017) Cellular and disease functions of the Prader-Willi Syndrome gene *MAGEL2*. *Biochem J* 474, 2177-90.
125. Tardivel P., J C, Canlet C., Lefort G., Tremblay-Franco M., Debrauwer L., Concordet D. & Servien R. (2017) ASICS: an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra. *Metabolomics*.
126. Tarrade A., Panchenko P., Junien C. & Gabory A. (2015) Placental contribution to nutritional programming of health and diseases: epigenetics and sexual dimorphism. *J Exp Biol* 218, 50-8.
127. Tasan M., Musso G., Hao T., Vidal M., MacRae C.A. & Roth F.P. (2015) Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat Methods* 12, 154-9.
128. Terenina E., Sautron V., Ydier C., Bazovkina D., Sevin-Pujol A., Gress L., Lippi Y., Naylies C., Billon Y., Liaubet L., Mormede P. & Villa-Vialaneix N. (2017) Time course study of the response to LPS targeting the pig immune gene networks. *BMC Genomics* 18, 988.
129. Tronchère H., Laporte J., Pendaries C., Chaussade C., Liaubet L., Pirola L., Mandel J.L. & Payrastré B. (2004) Production of phosphatidylinositol 5-phosphate by the phosphoinositide 3-phosphatase myotubularin in mammalian cells. *J Biol Chem* 279, 7304-12.
130. Turunen M., Olsson J. & Dallner G. (2004) Metabolism and function of coenzyme Q. *Biochim Biophys Acta* 1660, 171-99.
131. van der Lende T., Knol E.F. & Leenhouwers J.I. (2001) Prenatal development as a predisposing factor for perinatal losses in pigs. *Reprod Suppl* 58, 247-61.
132. Van Laere A.S., Nguyen M., Braunschweig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M. & Andersson L. (2003) A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* 425, 832-6.
133. Varrault A., Gueydan C., Delalbre A., Bellmann A., Houssami S., Aknin C., Severac D., Chotard L., Kahli M., Le Digarcher A., Pavlidis P. & Journot L. (2006) *Zac1* regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev Cell* 11, 711-22.
134. Velez-Irizarry D., Casiro S., Daza K.R., Bates R.O., Raney N.E., Steibel J.P. & Ernst C.W. (2019) Genetic control of longissimus dorsi muscle gene expression variation and joint analysis with phenotypic quantitative trait loci in pigs. *BMC Genomics* 20, 3.
135. Verbrugge S.A.J., Schönfelder M., Becker L., Yaghoob Nezhad F., Hrabě de Angelis M. & Wackerhage H. (2018) Genes Whose Gain or Loss-Of-Function Increases Skeletal Muscle Mass in Mice: A Systematic Literature Review. *Front Physiol* 9, 553.
136. Villa-Vialaneix N., Liaubet L., Laurent T., Cherel P., Gamot A. & SanCristobal M. (2013) The structure of a gene co-expression network reveals biological functions underlying eQTLs. *PLoS One* 8, e60045.
137. Villa-Vialaneix N., Liaubet L. & San Cristobal M. (2016) Depicting gene co-expression networks

- underlying eQTLs. In: *Systems Biology in Animal Production and Health* (ed. by H.N. Kadarmideen), pp. 1-31.
138. Villa-Vialaneix N., Liaubet L. & Sancristobal M. (2011) What is a good (gene) network? *J Anim Breed Genet* 128, 1-2.
 139. Voillet V., Besse P., Liaubet L., San Cristobal M. & González I. (2016) Handling missing rows in multi-omics data integration: multiple imputation in multiple factor analysis framework. *BMC Bioinformatics* 17, 402.
 140. Voillet V., San Cristobal M., Père M.C., Billon Y., Canario L., Liaubet L. & Lefaucheur L. (2018) Integrated Analysis of Proteomic and Transcriptomic Data Highlights Late Fetal Muscle Maturation Process. *Mol Cell Proteomics* 17, 672-93.
 141. Voillet V., SanCristobal M., Lippi Y., Martin P.G., Iannuccelli N., Lascor C., Vignoles F., Billon Y., Canario L. & Liaubet L. (2014) Muscle transcriptomic investigation of late fetal development identifies candidate genes for piglet maturity. *BMC Genomics* 15, 797.
 142. Wheeler T.L., Shackelford S.D. & Koohmaraie M. (2000) Variation in proteolysis, sarcomere length, collagen content, and tenderness among major pork muscles. *J Anim Sci* 78, 958-65.
 143. Workalemahu T., Grantz K.L., Grewal J., Zhang C., Louis G.M.B. & Tekola-Ayele F. (2018) Genetic and Environmental Influences on Fetal Growth Vary during Sensitive Periods in Pregnancy. *Sci Rep* 8, 7274.
 144. Yao Y., Voillet V., Jegou M., SanCristobal M., Dou S., Romé V., Lippi Y., Billon Y., Père M.C., Boudry G., Gress L., Iannucelli N., Mormède P., Quesnel H., Canario L., Liaubet L. & Le Huërou-Luron I. (2017) Comparing the intestinal transcriptome of Meishan and Large White piglets during late fetal development reveals genes involved in glucose and lipid metabolism and immunity as valuable clues of intestinal maturity. *BMC Genomics* 18, 647.
 145. Young F.G. (1957) Claude Bernard and the discovery of glycogen; a century of retrospect. *Br Med J* 1, 1431-7.
 146. Yvert G., Brem R.B., Whittle J., Akey J.M., Foss E., Smith E.N., Mackelprang R. & Kruglyak L. (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 35, 57-64.
 147. Zimmer D.B. & Magnuson M.A. (1990) Immunohistochemical localization of phosphoenolpyruvate carboxykinase in adult and developing mouse tissues. *J Histochem Cytochem* 38, 171-8.

Partie 2. Curriculum Vitae, Liste des publications, Activités d'enseignement et Responsabilités collectives

A. Curriculum Vitae

Laurence Liaubet
née le 18 janvier 1969
<http://genphyse.toulouse.inra.fr/groups/genorobust>

Tél. : 05 61 28 51 13
e-mail : laurence.liaubet@inra.fr

Fonction à l'INRA

Chargée de Recherche, recrutement octobre 2001 : Analyse du transcriptome musculaire du porc

Département : Génétique Animale

Unité : Génétique Physiologie et Systèmes d'élevage, UMR1388

Equipe : GENOROBUST Génétique des Systèmes en lien avec la Robustesse et l'Adaptation

Adresse : 24 chemin de Borde Rouge, 31326 Castanet Tolosan, France

Management

2016-...	Membre du comité de l'action COST CA15134 « GroupHouseNet », responsable du WG1 "Genetics and Epigenetics"
2014-...	Animatrice de l'équipe GENOROBUST
2013-2016	Responsable du WP « Transcriptomics » du projet ANR SusOStress, impliquée dans le WP « Data mining, biostatistics, and modelling »
2013-2016	Impliquée dans le WP « Transcriptomics » et « Integrative Biology » du projet ANR PigHeaT
2010-2014	Coordinatrice du projet ANR PORCINET

Formation

2000	Doctorat de l'Université Paul Sabatier-Toulouse III , Biologie du Développement, <i>mention Très Honorable avec les félicitations du jury.</i>
1996	DEA de Pharmacologie Moléculaire et Cellulaire , Toulouse
1995	Doctorat de Pharmacie , Toulouse, option industrie/recherche, <i>mention Très Honorable</i>
1993	Maîtrise des Sciences Biologiques et Médicales , Toulouse

Formations complémentaires

2015	MixOmics (14h) Omics Data Integration Project
2013-2014	<i>Management</i> (11 jours)
2011	Risques psychosociaux (8h)
2010	<i>Management</i> (2 jours)
2009	Etude du métabolome (1 semaine)
2007	Génomique des caractères complexes (1 semaine)
2004	Ecole du transcriptome (INSERM, 1 semaine, Marseille)
2004	Exploration statistique des données d'expression (INRA, 1 semaine)
2003	Introduction à l'analyse statistique (INRA, 1 semaine)
2002	De l'information bioinformatique de données de génomique fonctionnelle aux réseaux moléculaires (INSERM, 3 jours)
2001	Introduction aux bases de données moléculaires (CNRS/INSERM/INRA, 5 jours)
1998	Certificat de niveau 1 d'Utilisation et Protection des Animaux de Laboratoire
1994-1995	3 UV de Maîtrise , Université Toulouse III, 1-Génie génétique, 2-Outils de recherches en génétique et biologie moléculaire, 3-Biologie moléculaire végétale

Travaux scientifiques

- 2014-2019 **Chargée de Recherche dans l'équipe GENOROBUST pour Génétique des Systèmes en lien avec la Robustesse et l'Adaptation que j'anime**
Axe 1 : Déterminisme génétique de la variabilité de l'axe corticotrope, au centre des régulations des réponses de stress
Axe 2 : Déterminisme génétique et physiologique de la variabilité des caractères de production et d'adaptation (foie et/ou muscle)
Axe 3 : Impact de l'environnement précoce sur la programmation métabolique
- 2010-2013 **Chargée de Recherche dans l'équipe de Génétique Porcine animée par Juliette Riquet**
Génomique du muscle porcine en lien avec des phénotypes de robustesse (survie néonatale, réponses de stress, adaptation à la chaleur), de croissance et de production, intégration de données génétiques, transcriptome et métabolome
- 2002-2010 **Chargée de Recherche dans l'équipe GDO animée par François Hatey puis par Gwénola Tosser-Klopp**
Transcriptome du muscle porcine en lien avec des phénotypes de croissance et de production
- 2001 **Stage post-doctoral, INSERM U326, Toulouse/Purpan, dans l'équipe de Bernard Payrastra et UPR6520 CNRS/ULP/INSERM, Illkirch, dans l'équipe de Jean-Louis Mandel.**
Etude fonctionnelle de la myotubularine, une protéine déficiente dans les myopathies myotubulaires
- 1996-2000 **Thèse de l'Université Toulouse III, Centre de Biologie du Développement, CNRS UMR 5547, équipe de Fabienne Pituello.**
Contrôle de la régionalisation de la moelle épinière : identification et caractérisation d'une métalloprotéase
- 1995-1996 **DEA de Pharmacologie Moléculaire et Cellulaire** (2 stages de 5 mois), Toulouse
Institut de Pharmacologie et de Biologie Structurale, UMR 5089 CNRS, « Surexpression du récepteur μ aux opioïdes chez *S. cerevisiae* et par le système Sf9/baculovirus ».
Laboratoire de Génétique Cellulaire, UR444 INRA, « Analyse génétique et fonctionnelle de la folliculogénèse ovarienne de la truie ».
- 1995 **Thèse de Pharmacie réalisée dans le Laboratoire d'Interactions Plantes-Microorganismes UMR441 CNRS/INRA dans l'équipe d'Yves Marco**
Mise au point d'un modèle d'étude de l'interaction d'*A. thaliana* avec *R. solanacearum* et étude d'un gène codant un cytochrome P450 (réaction hypersensible)
- 1991-1994 Stages d'initiation à la recherche (Biotechnologie végétale, Pharmacologie moléculaire)

Enseignement - autres expériences professionnelles

- Novembre 2018 : Cours “*Omics approaches towards better understanding of the biology and sources of individual variation in stress responsiveness and behavior: finding biomarkers for trait and state*” avec Eduard Murani dans le cadre d’une *Training school* organisée par Laurianne Canario (Action COST GroupHouseNet)
- 2002-... Master 1 Bioingénierie Médicale (Université Paul Sabatier/Pharmacie), 2h/an, « La génomique fonctionnelle »
Master 2 Bioingénierie Biotechnologies Végétales (Université Paul Sabatier), 2h/an, « Introduction à la génomique Fonctionnelle »
Master 1 AgroFoodChain : 2h/an (en anglais), Functional Genomics, terminé en 2016
Master 2 Recherche Innovation Pharmacologique (Université Paul Sabatier) : 3h/an – atelier « Statistiques appliquées aux données transcriptomiques haut débit » avec M. San Cristobal, terminé
Formation statistiques données post-génomiques, niveau 3 (n’a pas eu lieu chaque année) : 1h, « Des graphes pour la biologie », cours également donné en tant qu’invitée à l’Université de Tours
- 2004 Gènes, Cellules et Développement : « Génomique fonctionnelle », cours master (M2R), environ 2h et analyses de publications
- 2001 **Qualifiée** aux fonctions de maître de conférences par trois sections du CNU (65, 85, 87)
- 2000 54 heures travaux pratiques Biologie du Développement, DEUG 2ème année
- 1999 2 heures cours, « Utilisation des levures en biotechnologie » certificat de Biotechnologie, Faculté de Pharmacie
- 1993-1994 **Stages hospitaliers** (1 an à mi-temps, 5ème année Pharmacie) : Cancérologie et Pharmacovigilance, Pharmacocinétique, Surveillance des infections nosocomiales.

B. Liste des publications

Travaux depuis mon recrutement à l’INRA (en bleu les étudiants en thèse que j’ai co-encadrés, en vert des étudiants en master, en thèse ou post-doc avec qui j’ai collaboré)

1. **Lefort** G, **Liaubet** L, Canlet C, Tardivel P, Père MC, Quesnel H, Paris A, Iannuccelli N, Vialaneix N, Servien R. ASICS: an R package for a whole analysis workflow of 1D 1H NMR spectra. *Bioinformatics*. 2019 Apr 12
2. **Marti-Marimon** M., Vialaneix N., Voillet V., Yerle-Bouissou M., Lahbib-Mansais Y. & **Liaubet** L. (2018) A new approach of gene co-expression network inference reveals significant biological processes involved in porcine muscle development in late gestation. *Sci Rep* 8, 10150.
3. Guerrero-Bosagna C, Morisson M, **Liaubet** L, Rodenburg TB, de Haas EN, Košťál Ľ, Pitel F. Transgenerational epigenetic inheritance in birds. *Environ Epigenet*. 2018 Apr 26;4(2):dvy008.
4. Gondret F, Guével B, Père MC, Quesnel H, Billon Y, Com E, Canario L, Louveau I, **Liaubet** L. Proteomic analysis of adipose tissue during the last weeks of gestation in pure and crossbred Large White or Meishan fetuses gestated by sows of either breed. *J Anim Sci Biotechnol*. 2018 Apr 3;9:28.
5. **Voillet** V, San Cristobal M, Père MC, Billon Y, Canario L, **Liaubet** L, Lefaucheur L. Integrated Analysis of Proteomic and Transcriptomic Data Highlights Late Fetal Muscle Maturation Process. *Mol Cell Proteomics*. 2018 Apr;17(4):672-693.
6. Terenina E, **Sautron** V, **Ydier** C, Bazovkina D, Sevin-Pujol A, Gress L, Lippi Y, Naylies C, Billon Y, **Liaubet** L, Mormede P, Villa-Vialaneix N. Time course study of the response to LPS targeting the pig immune gene networks. *BMC Genomics*. 2017 Dec 22;18(1):988.
7. **Dou** S, Villa-Vialaneix N, **Liaubet** L, Billon Y, Giorgi M, Gilbert H, Gourdine JL, Riquet J, Renaudeau D. 1HNMR-Based metabolomic profiling method to develop plasma biomarkers for sensitivity to chronic heat stress in growing pigs. *PLoS One*. 2017 Nov 27;12(11):e0188469.

8. **Yao Y, Voillet V, Jegou M, SanCristobal M, Dou S, Romé V, Lippi Y, Billon Y, Père MC, Boudry G, Gress L, Iannuccelli N, Mormède P, Quesnel H, Canario L, Liaubet L, Le Huërou-Luron I.** Comparing the intestinal transcriptome of Meishan and Large White piglets during late fetal development reveals genes involved in glucose and lipid metabolism and immunity as valuable clues of intestinal maturity. *BMC Genomics*. 2017 Aug 22;18(1):647.
9. **François Y, Vignal A, Molette C, Marty-Gasset N, Davail S, Liaubet L, Marie-Etancelin C.** Deciphering mechanisms underlying the genetic variation of general production and liver quality traits in the overfed mule duck by pQTL analyses. *Genet Sel Evol*. 2017 Apr 19;49(1):38.
10. **Voillet V, Besse P, Liaubet L, San Cristobal M, González I.** Handling missing rows in multi-omics data integration: multiple imputation in multiple factor analysis framework. *BMC Bioinformatics*. 2016 Oct 3;17(1):402
11. SanCristobal M, **Rohart F, Lascor C, Bouffaud M, Trouilh L, Martin PG, Lippi Y, Tribout T, Faraut T, Mercat MJ, Milan D, Liaubet L.** Exploring transcriptomic diversity in muscle revealed that cellular signaling pathways mainly differentiate five Western porcine breeds. *BMC Genomics*. 2015 Dec 12;16(1):1055.
12. **Sautron V, Terenina E, Gress L, Lippi Y, Billon Y, Larzul C, Liaubet L, Villa-Vialaneix N, Mormède P.** Time course of the response to ACTH in pig: biological and transcriptomic study. *BMC Genomics*. 2015 Nov 17;16(1):961.
13. Pierron A, Mimoun S, Murate LS, Loiseau N, Lippi Y, Bracarense AF, **Liaubet L, Schatzmayr G, Berthiller F, Moll WD, Oswald IP.** Intestinal toxicity of the masked mycotoxin deoxynivalenol-3- β -D-glucoside. *Arch Toxicol*. 2015 Sep 24.
14. Picard B, Lebret B, Cassar-Malek I, **Liaubet L, Berri C, Le Bihan-Duval E, Hocquette JF, Renand G.** Recent advances in omic technologies for meat quality management. *Meat Sci*. 2015 Nov;109:18-26. *Revue*
15. **Congras A, Yerle-Bouissou M, Pinton A, Vignoles F, Liaubet L, Ferchaud S, Acloque H.** Sperm DNA methylation analysis in swine reveals conserved and species-specific methylation patterns and highlights an altered methylation at the GNAS locus in infertile boars. *Biol Reprod*. 2014 Dec;91(6):137.
16. **Voillet V, SanCristobal M, Lippi Y, Martin PG, Iannuccelli N, Lascor C, Vignoles F, Billon Y, Canario L, Liaubet L.** Muscle transcriptomic investigation of late fetal development identifies candidate genes for piglet maturity. *BMC Genomics*. 2014 Sep 17;15:797.
17. Villa-Vialaneix N, **Liaubet L, Laurent T, ChereL P, Gamot A, SanCristobal M.** The structure of a gene co-expression network reveals biological functions underlying eQTLs. *PLoS One*. 2013;8(4):e60045.
18. **Rohart F, Paris A, Laurent B, Canlet C, Molina J, Mercat MJ, Tribout T, Muller N, Iannuccelli N, Villa-Vialaneix N, Liaubet L, Milan D, San Cristobal M.** Phenotypic prediction based on metabolomic data for growing pigs from three main European breeds. *J Anim Sci*. 2012 Dec;90(13):4729-40.
19. **Lacomme M, Liaubet L, Pituello F, Bel-Vialar S.** NEUROG2 drives cell cycle exit of neuronal precursors by specifically repressing a subset of cyclins acting at the G1 and S phases of the cell cycle. *Mol Cell Biol*. 2012 Jul;32(13):2596-607.
20. **Liaubet L, Lobjois V, Faraut T, Tircazes A, Benne F, Iannuccelli N, Pires J, Glénisson J, Robic A, Le Roy P, Sancristobal M, ChereL P.** Genetic variability of transcript abundance in pig perimortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. *BMC Genomics*. 2011 Nov 4;12:548.
21. Robic A, Fève K, Larzul C, Billon Y, van Son M, **Liaubet L, Sarry J, Milan D, Grindflek E, Bidanel JP, Riquet J.** Expression levels of 25 genes in liver and testis located in a QTL region for androstrenone on SSC7q1.2. *Anim Genet*. 2011 Dec;42(6):662-5
22. Bonnet A, Bevilacqua C, Benne F, Bodin L, Cotinot C, **Liaubet L, Sancristobal M, Sarry J, Terenina E, Martin P, Tosser-Klopp G, Mandon-Pepin B.** Transcriptome profiling of sheep granulosa cells and oocytes during early follicular development obtained by laser capture microdissection. *BMC Genomics*. 2011 Aug 18;12:417.
23. Villa-Vialaneix N, **Liaubet L, Sancristobal M.** What is a good (gene) network? *J Anim Breed Genet*. 2011 Feb;128(1):1-2. *Editorial*

24. Jouffe* V, Rowe* S, **Liaubet*** L, Buitenhuis B, Hornshøj H, Sancristobal M, Mormède P, de Koning DJ. Using microarrays to identify positional candidate genes for QTL: the case study of ACTH response in pigs. *BMC Proc.* 2009 Jul 16;3 Suppl 4:S14. (*co-1er auteurs)
25. Hedegaard J, Arce C, Biciato S, Bonnet A, Buitenhuis B, Collado-Romero M, Conley LN, Sancristobal M, Ferrari F, Garrido JJ, Groenen MA, Hornshøj H, Hulsegge I, Jiang L, Jiménez-Marín A, Kommadath A, Lagarrigue S, Leunissen JA, **Liaubet** L, Neerincx PB, Nie H, van der Poel J, Prickett D, Ramirez-Boo M, Rebel JM, Robert-Granié C, Skarman A, Smits MA, Sørensen P, Tosser-Klopp G, Watson M. Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. *BMC Proc.* 2009 Jul 16;3 Suppl 4:S5.
26. Bonnet A, Lagarrigue S, **Liaubet** L, Robert-Granié C, Sancristobal M, Tosser-Klopp G. Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares. *BMC Proc.* 2009 Jul 16;3 Suppl 4:S11.
27. Robic A, Faraut T, **Liaubet** L, Milan D. The carnitine acetyltransferase gene (CRAT): a characterization of porcine transcripts with insights into the 5'-end variants of mammalian transcripts and their possible sub-cellular localization. *Cell Mol Biol Lett.* 2009;14(1):90-9.
28. Robic A, Faraut T, **Liaubet** L, Riquet J, Milan D, Lobjois V. Characterization of porcine ASB6 gene and transcripts-comparison of mammalian transcripts. *Anim Biotechnol.* 2008;19(3):138-43
29. **Lobjois** V*, **Liaubet** L*, SanCristobal M, Glénisson J, Fève K, Rallières J, Le Roy P, Milan D, Cherel P, Hatey F. A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. *Anim Genet.* 2008 Apr;39(2):147-62. (*co-1er auteurs)
30. **Hazard** D, **Liaubet** L, Sancristobal M, Mormède P. Gene array and real time PCR analysis of the adrenal sensitivity to adrenocorticotropic hormone in pig. *BMC Genomics.* 2008 Feb 27;9:101.
31. **Ferré** PJ*, **Liaubet** L*, Concordet D, SanCristobal M, Uro-Coste E, Tosser-Klopp G, Bonnet A, Toutain PL, Hatey F, Lefebvre HP. Longitudinal analysis of gene expression in porcine skeletal muscle after post-injection local injury. *Pharm Res.* 2007 Aug;24(8):1480-9. (*co-1er auteurs)
32. **Demeure** O, **Liaubet** L, Riquet J, Milan D. Determination of PRKAG1 coding sequence and mapping of PRKAG1 and PRKAG2 relatively to porcine back fat thickness QTL. *Anim Genet.* 2004 Apr;35(2):123-5. PubMed PMID: 15025572.

Travaux post-doctoraux

33. Tronchère H, Laporte J, Pendaries C, Chaussade C, **Liaubet** L, Pirola L, Mandel JL, Payrastré B. Production of phosphatidylinositol 5-phosphate by the phosphoinositide 3-phosphatase myotubularin in mammalian cells. *J Biol Chem.* 2004 Feb 20;279(8):7304-12.
34. Laporte J*, **Liaubet** L*, Blondeau F*, Tronchère H, Mandel JL, Payrastré B. Functional redundancy in the myotubularin family. *Biochem Biophys Res Commun.* 2002 Feb 22;291(2):305-12. (*co-1er auteurs)

Travaux de ma thèse en Sciences

35. **Liaubet** L, Bertrand N, Medevielle F, Pituello F. Identification by differential display of a chicken tolloid-related metalloprotease specifically expressed in the caudal notochord. *Mech Dev.* 2000 Aug;96(1):101-5.

Travaux de DEA

36. Gaibelet G, Meilhoc E, Rioud J, Saves I, Exner T, **Liaubet** L, Nürnberg B, Masson JM, Emorine LJ. Nonselective coupling of the human mu-opioid receptor to multiple inhibitory G-protein isoforms. *Eur J Biochem.* 1999 Apr;261(2):517-23.

Travaux de ma thèse en Sciences Pharmaceutiques

37. Godiard L, Sauviac L, Dalbin N, **Liaubet L**, Callard D, Czernic P, Marco Y. CYP76C2, an Arabidopsis thaliana cytochrome P450 gene expressed during hypersensitive and developmental cell death. FEBS Lett. 1998 Nov 6;438(3):245-9.
38. Deslandes L, Pileur F, **Liaubet L**, Camut S, Can C, Williams K, Holub E, Beynon J, Arlat M, Marco Y. Genetic characterization of RRS1, a recessive locus in Arabidopsis thaliana that confers resistance to the bacterial soilborne pathogen Ralstonia solanacearum. Mol Plant Microbe Interact. 1998 Jul;11(7):659-67.

Rapports diplômants

Liaubet L. Mise au point d'un modèle d'étude de l'interaction d'A. thaliana avec R. solanacearum et étude d'un gène codant un cytochrome P450 (réaction hypersensible) ». Thèse pour obtenir le grade de Docteur en Pharmacie de l'Université Paul Sabatier-Toulouse III. Septembre 1995.

Liaubet L. Contrôle de la régionalisation de la moelle épinière : identification et caractérisation d'une métalloprotéase spécifiquement exprimée dans la région caudale de la chorde. Thèse pour obtenir le grade de Docteur de l'Université Paul Sabatier-Toulouse III. Décembre 2000.

C. Encadrements

La très grande majorité des étudiants que j'encadre sont des étudiants en statistique et beaucoup moins en bio-informatique et en biologie. Ceci souligne que l'investissement nécessaire à la réalisation d'études mettant en jeu des données de grandes dimensions et d'origines diverses nécessitent des compétences qui s'éloignent de la biologie cellulaire et moléculaire (mon socle de connaissances). Autre particularité, pour assurer l'encadrement, les étudiants sont co-encadrés avec un statisticien ou bien un bio-informaticien.

Etudiants de divers niveaux

- Sandrine Noguera, stage BTS Biochimie 1ère année, juin 2002.
- Marie Duchamp, stage M1R, MSBM de Biotechnologie et Ingénierie Biomédicale, juin 2004. Hybridation sur microréseaux.
- Vincent Cochard, stage 3ème année de l'ESAP (Ecole Supérieure d'Agriculture de Purpan), 10 semaines entre juillet et octobre 2005. PCR en temps réel.
- Aurélie Tircazes. Etude de l'expression des gènes d'un muscle strié squelettique porcin, Longissimus dorsi par l'analyse du transcriptome : obtention et analyse des ARN. BTS Anabiotech par alternance. Septembre 2004-août 2006.
- Carole Arfelis, M1 Informatique Statistique Mathématique Appliqués à la Gestion, Avril-Juillet 2007. Co-encadrée par M SanCristobal)
- Jérôme Bertrandias, DESS Statistique et Econométrie 2004, co-encadré par M SanCristobal
- Fanny Guimaraes, 4ème année INSA, Juillet Août 2007, co-encadrée par C Robert-Granié.
- M. Vermerie. Etude de la tendreté de la longe de porc par une analyse transcriptomique : localisation chromosomique de sept gènes d'intérêt. DUT Bioinformatique. Juin 2006.
- C. Arfelis. Identification de gènes influençant la qualité de la viande de porc. M1 ISMAG. Juin 2007.
- J. Thia Ah Fat. Recherche de gènes influençant la qualité de la viande porcine. M1 Ingénierie Mathématique. Co-encadré par M. San Cristobal. Mai-juillet 2008.
- C. Lascor. Etude du transcriptome du muscle porcin : analyse de gènes potentiellement prédicteurs de la tendreté de la viande par PCR en temps réel à haut débit. BTS Anabiotech par alternance. Septembre 2008-août 2010.
- S. Parisis. Contribution à l'annotation du génome du porc par génomique comparative et amélioration de l'annotation de 40000 transcrits porcins. Licence Professionnelle Bioinformatique. Septembre 2010. Co-encadré par Thomas Faraut.

- A. Poujol. Influence de l'anesthésie sur les paramètres physiologiques des truies et des fœtus : Suivi de 39 césariennes. Thèse Vétérinaire (Ecole Nationale Vétérinaire de Toulouse) 2011.
- Cahuzac L. Diagne F. et Verger M. Analyse statistique de données transcriptomiques haut débit : étude du développement fœtal et recherche de marqueurs de maturité. Projet tutoré, Génie Mathématiques et Modélisation 4eme année, INSA, janvier à mai 2013. Co-encadrés par M. San Cristobal.
- Diagne F. Analyse du transcriptome du sang total, recherche de prédicteurs et intégration de données multiples. Co-encadrée par M. San Cristobal, Stage de Génie Mathématiques et Modélisation 4eme année 4ème année INSA, juin à septembre 2013.
- Mathieu K. Amélioration de l'annotation des 60000 séquences d'oligo-nucléotides (60-mers) d'une puce d'expression génique. Stage Licence Pro Bioinformatique, IUT de Perpignan – Département STID, 2014. Co-encadré par C. Klopp.
- Platon J et Bousquet F. Analyse et intégration de données transcriptome pour modéliser la fin du développement fœtal. Projet tutoré, 4ème année Génie Mathématique et Modélisation, INSA, janvier à mai 2014. Co-encadrés par V. Voillet.
- Castaing Y et Khalil L. Analyse et intégration de données transcriptomes afin de modéliser la fin du développement fœtal. Projet tutoré, 4ème année Génie Mathématique et Modélisation, INSA, janvier à mai 2015. Co-encadrés par V. Voillet.
- Champion C. Analyse des données métabolomiques de PigHeaT (adaptation à la chaleur). Stage M1 Mathématiques Appliquées Pour l'Ingénierie, l'Industrie et l'Innovation (MAPI3), mai-juillet 2016. Co-encadrée par M. San Cristobal et N. Vialaneix.

Etudiants niveau master 2

- Gamot A. Analyses de grands réseaux biologiques : construction et structure. M2 Statistique et Econométrie. Co-encadré par N. Vialaneix et M. San Cristobal, mars-août 2009.
- Yammine S. Etude de la fin du développement foetal en relation avec la survie à la naissance, transcriptome surrénalien et biologie de l'axe corticotrope. M2R EQSA (Elaboration de la Qualité et de la Sécurité Alimentaire), ENSAT. Co-encadré par E. Terenina, janvier-juin 2012.
- Voillet V. Analyse et intégration de données transcriptomiques pour modéliser la fin du développement fœtal. M2R Bioinformatique et Biologie des Systèmes. Université Paul Sabatier TOULOUSE III. Co-encadré par M. San Cristobal, janvier-juin 2013.
- Ydier C. Etude de la réponse au lipopolysaccharide chez le porcelet. M2 Ingénierie Mathématique. Co-encadrée par N. Vialaneix et V. Sautron, avril-septembre 2015.
- Salanon T. Etude de la survie néonatale chez le porcelet. Stage de fin d'étude de l'école d'Ingenieurs de Purpan, 6 mois entre juin et décembre 2017.
- Guilmineau C. Analyse des données métabolomiques du projet SubPig. Master 2 Ingénierie de la décision et big data à l'Institut de Mathématiques Appliquées, Université Catholique de l'Ouest, Angers. Co-encadrée par N. Vialaneix, Rémi Servien et Gaëlle Lefort, Fev-Juil 2019.
- Mercier L. Analyse et intégration de données métabolomiques pour modéliser la réponse de quatre lignées génétiques à un challenge alimentaire. Mémoire Fin d'Etudes, Spécialité M2 Mathématiques et applications. Université de Strasbourg, France. Co-encadrée par H. Gilbert et G. Lefort, en collaboration avec Mette Skou Hedemann (Danemark). Mars-Août 2019.

Etudiants en thèse

Autre particularité de mon parcours, il n'y a qu'un seul étudiant en thèse, Valentin Voillet, que j'ai co-encadré de façon officielle avec Magali San Cristobal. Je co-encadre maintenant une deuxième étudiante en thèse, Gaëlle Lefort.

Les autres étudiants ont réalisé leur thèse en biologie cellulaire et moléculaire, en génétique moléculaire, en pharmacologie, ou bien encore en toxicologie, et selon l'avancement de leurs travaux, ils trouvent un intérêt à réaliser une étude transcriptomique. Je participe alors à l'encadrement de l'étude en question

avec un accompagnement du plan d'expérience à l'interprétation des données. Les noms en gras sont ceux des étudiants en thèse pour lesquels je considère mon investissement comme étant significatif.

- Depuis septembre 2018, je participe à l'encadrement de la thèse de **Gaëlle Lefort** avec Nathalie Vialaneix (Unité MIA-T de l'INRA) et Rémi Servien (Unité Intheres de l'INRA) ainsi que Hélène Quesnel (Unité PEGASE de l'INRA). Financement #DigitAg 50%, 50% départements INRA (GA, SA et MIA). Le sujet est l'analyse des données métabolomiques (Lefort et al. 2019b) et l'intégration des données du projet PORCINET et SuBPig.
- Océane Schmitt, étudiante en thèse en Irlande soutenue en 2018, "Investigating management strategies of large litters in pigs". Université d'Edimbourg, Roslin Institute et Teagasc Encadrantes : Emma Baxter, Keelin O'Driscoll and Laura Boyle. Venue 1 mois en janvier 2018 dans le cadre d'une mission courte, STSM du COST GroupHouseNet, thermographie à la naissance.
- Maria Marti-Marimon, Master puis Thèse « 3D genome conformation and gene expression in fetal pig muscle at late gestation » (Prix Léopold Escande) réalisée au laboratoire GenPhySE, encadrée par Martine Yerle et Sylvain Foissac, soutenue en 2018 (Marti-Marimon *et al.* 2018).
- **Valentin Voillet**. Thèse soutenue en 2016 (Prix Léopold Escande). La thèse « PorciMuscle » est consacrée à l'analyse et l'intégration des données de PORCINET ; intégration au niveau du tissu musculaire (transcriptome, protéome, typage des myosines et glycogène) et intégration au niveau de l'animal en prenant en compte toutes les informations jugées pertinentes afin de modéliser la mise en place de la maturité (Voillet *et al.* 2014; Voillet *et al.* 2016; Voillet *et al.* 2018). Financement 25% département GA, 25% département PHASE, 50% région Midi-Pyrénées. Co-encadré par M. San Cristobal et en collaboration avec Louis Lefaucheur à St Gilles (UMR 1348 PEGASE, INRA). Sept 2013- Oct 2016
- Valérie Sautron. Thèse sur l'analyse et intégration des données obtenues dans le cadre d'un projet ANR SusOStress, Génétique des réponses de stress chez le porc (Sautron et al. 2015; Terenina et al. 2017). Encadrée par E. Terenina, N. Villa-Vialaneix et P. Mormède. Oct 2013- Sept 2016. Je participe au suivi des analyses des données transcriptomiques (Sautron *et al.* 2015; Terenina *et al.* 2017) pour cette thèse qui se déroule dans mon équipe et je suis la responsable de la tâche transcriptome du projet ANR SusOStress.
- Annabelle Congras. Analyse de la méthylation de l'ADN spermatique et développement de cellules pluripotentes induites chez des verrats infertiles porteurs ou non de remaniements chromosomiques. Thèse soutenue en 2014. Cette thèse inclut une analyse du transcriptome de cellule iPS et a nécessité l'utilisation d'une puce porcine 60K (Congras et al. 2014). Encadrée par Martine Yerle et Hervé Acloque.
- Florian Rohart. Thèse soutenue en 2012 « Prédiction phénotypique et sélection de variables en grande dimension dans les modèles linéaires et linéaires mixtes » encadré par Magali SanCristobal et Béatrice Laurent. Il s'agissait d'analyser des données métabolomiques obtenues dans le cadre du projet sur la diversité fonctionnelle entre différentes espèces et de prédire les phénotypes de production (Rohart *et al.* 2012; SanCristobal *et al.* 2015).
- Marine Lacomme. Thèse soutenue en 2011. Réseaux géniques initiés par *NEUROG2* dans la moelle épinière en développement : analyse globale et étude de la régulation de *PAX6*. Cette thèse inclut une analyse du transcriptome (Lacomme et al. 2012). Encadrée par Sophie Bel-Vialar et Fabienne Pituello, Laboratoire de Biologie du Développement, UMR5547 CNRS, Université Toulouse III.
- Stefania Iacuniello. Etudiante en thèse italienne qui a bénéficié d'un financement COST (Action 86) pour venir de mars à juillet 2007, Directeur de thèse : Giulio Pagnacco, Professor of animal breeding, University of Milan, Italy. Projet sur qualité de la viande. Elle a effectué des analyses d'expression de gènes musculaires par qPCR et des localisations sur le panel RH (Gorni et al. 2008; Gorni et al. 2011). Je ne suis pas sûre qu'elle ait soutenu sa thèse et je n'ai pas été associée aux publications relatives aux travaux effectués dans notre laboratoire.
- Oliver Demeure. Thèse soutenue en 2004. Cartographie fine et caractérisation d'un QTL localise sur le chromosome 7 porcine et affectant les caractères de croissance et d'engraissement. Encadré par Denis Milan. Participation à la caractérisation moléculaire d'un gène candidat PRKAG1 (Demeure et al. 2004).

- **Pierre Ferré.** Thèse soutenue en 2004 (Prix Léopold Escande). Étude des lésions musculaires iatrogènes : méthodes non invasives d'évaluation quantitative et mécanismes physiopathologiques. Encadré par Hervé Lefebvre, UMR 181 INRA, ENVT. J'ai encadré Pierre Ferré pendant sa dernière année de thèse, il a été recruté chez Pierre Fabre avant que les analyses statistiques et l'interprétation des données soient terminées. J'ai donc assuré l'exploitation et la valorisation des données obtenues (Ferre *et al.* 2007).

Encadrement de jeunes docteurs

- Valérie Lobjois (Lobjois *et al.* 2008; Robic *et al.* 2008; Liaubet *et al.* 2011b). J'avais obtenu une bourse post-doctorale de l'INRA de 18 mois pour effectuer des analyses du transcriptome du muscle porcin. J'ai effectué la valorisation des données ; co-localisation avec des QTL, réseaux de gènes, écriture et soumission de la publication.
- Dominique Hazard (Hazard *et al.* 2008) était en stage post-doctoral avec Pierre Mormède qui avait souhaité que je l'encadre pour la réalisation d'une étude transcriptomique des glandes surrénales de porcs Large White ou Meishan soumis ou non à une injection d'ACTH. 1ères implications pour moi dans l'étude du rôle central de l'activité corticotrope dans la régulation des caractères de production et de robustesse.

D. Responsabilités collectives

1. Fonctionnement des collectifs

Au sein du laboratoire

Agent Chargé de Prévention de 2004 à 2013 et participant à ce titre au du comité de gestion du laboratoire de Génétique Cellulaire.

Participation aux réunions du Comité de Direction de GenPhySE depuis 2014 en tant qu'animatrice de l'équipe GENOROBUST.

Au sein de l'INRA

Groupe Biopuces animé d'abord par M. San Cristobal puis par N. Vialaneix et moi-même depuis septembre 2015. Des réunions mensuelles permettent des discussions très ouvertes entre biologistes et statisticiens autour de l'analyse de données de grande dimension (microarray, RNA-seq, métabolome, protéome, métagénome, méthylome...). Les thématiques abordées évoluent avec les technologies et les méthodes développées pour l'intégration des données. Ce groupe existe depuis plus de 10 ans et réunit des personnes INRA mais aussi INSERM, UPS et CNRS.

Sur le site toulousain pour diverses animations scientifiques

Membre du comité de plateforme Centre de Ressources Génotypage Séquençage de la génopole de Toulouse 2004 à 2010.

Membre depuis 2010 du comité plateforme Biostatistiques de Toulouse animée par S. Déjean (<http://perso.math.univ-toulouse.fr/biostat/presentation/lequipe/>)

Réseaux européens

Contribution à 2 groupes de travail européens « **EADGENE** », *European Animal Disease Génomique*

Network of Excellence for Animal Health and Food Safety (2006 et 2008). Ce groupe de travail avait pour objectif de proposer à différents groupes européens de travailler sur un même jeu de données pour confronter nos approches statistiques et post-génomiques. Les travaux auxquels j'ai participé ont donné lieu à trois publications (Bonnet et al. 2009; Hedegaard et al. 2009; Jouffe et al. 2009)

Participation à un réseau européen COST **GroupHouseNet**, *Synergy for preventing damaging behaviour in group housed pigs and chickens*, co-responsable du WG1 "Genetics and Epigenetics". Ce réseau a été l'occasion de recevoir une étudiante en thèse en Irlande pour analyser la cinétique de la 1^{ère} heure après la naissance (4 temps : 8, 15, 30, 60 min) par thermographique infrarouge des nouveau-nés. Cette étude permet de rajouter un lot de données phénotypiques (14 phénotypes) de différentes températures corporelles à nos phénotypes de maturité-croissance-mobilité et nos données d'expression et métabolomiques. Laurianne Canario a aussi organisé, dans le cadre de ce réseau COST, une *Training school* sur la génétique du comportement particulièrement intéressante au cours de laquelle j'ai donné un cours de génomique fonctionnelle avec Eduard Murani. Une revue est sortie de ce groupe de travail sur l'épigénétique (Guerrero-Bosagna et al. 2018) et une autre est en préparation sur les approches – omiques dans les études du comportement.

2. Animation de l'équipe GENOROBUST, Génétique des systèmes en lien avec l'adaptation et la Robustesse

Approche système

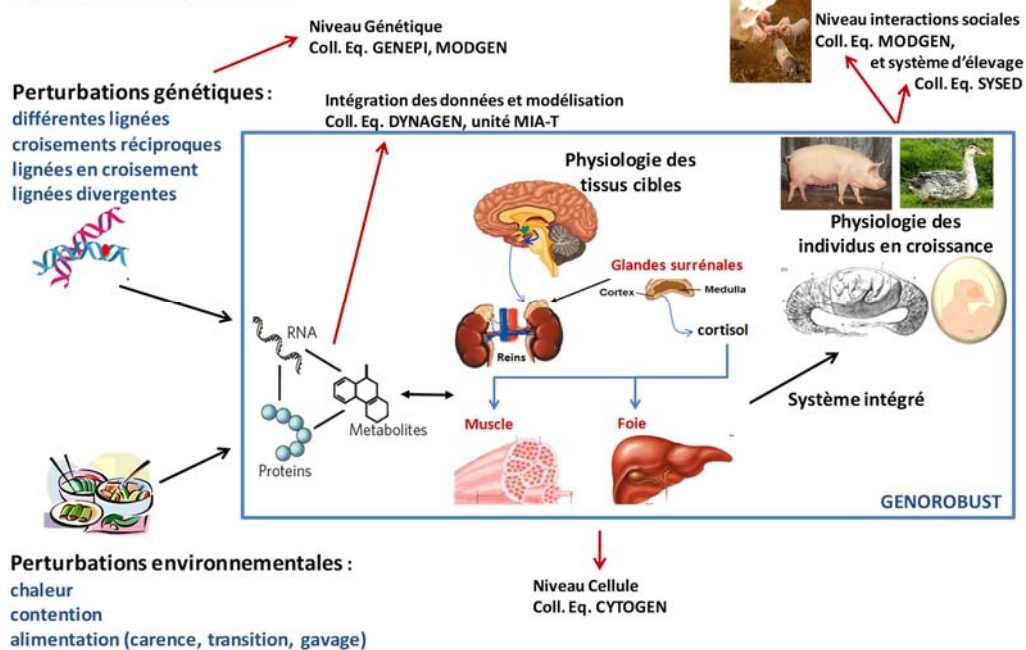


Figure 50 : Le projet de l'équipe Genorobust s'inscrit dans le cadre d'une approche système où les niveaux d'études de tissus, organes avec des études physiologiques, d'expression, du protéome et du métabolome. Dans l'encart bleu, ce que nous faisons dans l'équipe et en dehors nos collaborations proximales (GenPhySE et MIA-T). Pour modéliser le système, il faut introduire des perturbations telles que suggérées, perturbations génétiques et/ou perturbations environnementales (chaleur, stress, alimentaire).

En janvier 2014, le LGC (Laboratoire de Génétique Cellulaire, département Génétique Animale) a fusionné avec deux autres laboratoires, la SAGA (Station d'Amélioration Génétique des Animaux, département GA) et TANDEM (Tissus Animaux Nutrition Digestion Ecosystème et Métabolisme, département PHASE, Physiologie Animale et Système d'Élevage) pour constituer le laboratoire GenPhySE. Dans ce cadre, je fais partie d'une nouvelle équipe qui s'intéresse au déterminisme des caractères complexes chez le porc et le canard avec une approche intégrative (génomique et physiologie) ; équipe GENOROBUST. L'objectif est l'étude des composantes moléculaires (génétique, épigénétique, métaboliques) et la modélisation des compromis entre caractères de production et

d'adaptation. Cette équipe comprend en juillet 2019, 7 scientifiques, 5 ingénieurs et techniciens, 2 étudiants en thèse et j'en assure l'animation.

Notre projet revient à s'intéresser à l'équilibre à rechercher entre caractères de production et de robustesse, cet équilibre est à mettre en relation avec la théorie de l'allocation des ressources ; il s'agit donc d'étudier la régulation des métabolismes en lien avec la robustesse. Le projet s'articule autour des 3 axes suivants et sont résumés dans la Figure 50.

Axe 1 : Déterminisme génétique de la variabilité de l'axe corticotrope, au centre des régulations des réponses de stress

Axe 2 : Déterminisme génétique et physiologique de la variabilité des caractères de production et d'adaptation (foie et/ou muscle)

Axe 3 : Impact de l'environnement précoce sur la programmation métabolique et lien avec les caractères de production et d'adaptation

Ce projet est celui en place depuis 2014. Comme le laboratoire va être évalué en novembre 2019, nous avons retravaillé le projet d'équipe pour nous recentrer sur l'étude du développement précoce des phénotypes et la recherche d'indicateurs pour améliorer la robustesse. Notre approche continuera à utiliser des outils de biologie intégrative pour analyser des données multi-omiques et multi-tissus, et nous avons structuré notre projet en trois nouveaux axes.

Axe 1 : Phénotypage fin et variabilité individuelle. Cet axe a pour objectifs de terminer les études en cours et en particulier les travaux d'intégration des données générées dans les études réalisées ces 5 dernières années.

Axe 2 : Effet de l'environnement périnatal avec l'analyse de la programmation métabolique et des conditions de reproduction. C'est un axe très ambitieux pour lequel nous avons beaucoup de difficultés à financer des études très fondamentales. Voir dans ce document la partie sur les interactions fœto-maternelles.

Axe 3 : Recherche et validation des indicateurs. Cet axe correspond à notre objectif finalisé dont j'ai donné un exemple dans mes perspectives (survie néonatale et autres phénotypes) et nous en avons un aussi pour l'espèce canard avec la prédiction de caractères de production.