Modifications de la réponse transcriptomique au repas et des profils de méthylation de l'ADN dans le duodénum du porc par une sélection divergente sur l'efficacité alimentaire.

Safia Saci

SACI Safia

Du 1er janvier au 30 juin 2021

Rapport de stage

Formation :
Master 2 bio-informatique et biologie des systèmes
Université Toulouse 3 Paul Sabatier

Equipe d'accueil
Génétique et épigénétique (GenEpi)

Unité d'accueil
Génétique et physiologie des systèmes d'élevage, (genphyse) - INRAE Auzeville-Tolosane

Tuteur en entreprise
DEVAILLY Guillaume

Tuteur académique
FICHANT Gwenaëlle

Rapporteur
CLOUAIRE Thomas

Résumé

Améliorer l'efficacité alimentaire des animaux d'élevage permet de subvenir aux besoins nutritionnels de la population humaine, réduire les impacts environnementaux de l'élevage tout en réduisant les principaux coûts pour les éleveurs : l'achat d'aliments et exploitation des terres agricoles. Pour mieux comprendre l'efficacité alimentaire chez le porc, des lignées divergentes ont été établies, à partir d'un noyau de porcs de race Large White, sur le critère de Consommation Moyenne Journalière Résiduelle (CMJR), donnant, après 11 générations de sélection, une lignée de porcs dits efficaces (CMJR-) ou peu efficaces (CMJR+).

Durant ce projet, nous avons analysé des données complexes de séquençages haut débits concernant la méthylation et la transcription de l'ADN des deux lignées divergeant sur un critère d'efficacité alimentaire. Nous avons caractérisé le transcriptome du duodénum de ces deux lignées avant et après la prise alimentaire (n = 24). La réponse transcriptomique du duodénum à la prise alimentaire est plus importante que la différence de transcriptome entre les deux lignées de porcs. Les animaux CMJR- ont une réponse transcriptomique plus importante au passage du bol alimentaire que les animaux CMJR+. L'analyse de la méthylation de l'ADN à l'échelle du génome dans ces échantillons issus par précipitation d'ADN méthylés ont montré que le méthylome n'est pratiquement pas modifié par la prise alimentaire, mais les profils de méthylations du duodénum sont distincts entre les deux lignées.

Ainsi, la forte réponse transcriptomique du duodénum suite à la prise alimentaire n'a pas d'impact sur le méthylome et ne semble pas liée à des changements de méthylation de l'ADN à l'échelle d'un repas. La sélection sur le critère CMJR a modifié la réponse transcriptomique du duodénum au passage du bol alimentaire, et a entraîné des différences de méthylation de l'ADN entre ces deux lignées.

Abstract

Improving the feed efficiency of farm animals makes it possible to meet the nutritional needs of the population, reduce the environmental impacts of breeding while reducing the main costs for breeders: the purchase of animal feed and use of agricultural land. To understand the feed efficiency in pigs, divergent lines were established from a nucleus of large white pigs on the criterion of average daily residual consumption (CMJR), giving, after 11 generations of selection, an effective line. (CMJR-) or inefficient (CMJR). +) pigs.

During this project, we analyzed complex high-throughput sequencing data concerning methylation and DNA transcription of the two lines diverging on a criterion of feed efficiency. We characterized the duodenal transcriptome of these two lines before and after food intake (n = 24). The transcriptome response of the duodenum to food intake is greater than the difference in transcriptome between the two lines of pigs. CMJR- animals have a greater transcriptomic response to the passage of the food bolus than CMJR + animals. Methylation analysis of samples obtained by precipitation of methylated DNA at the genome-wide scale showed that the methylome was not affected by food intake, but the methylation profiles of the duodenum were distinct between the two lines.

Thus, the strong transcriptomic response of the duodenum following food intake has no impact and is not related to meal-scale changes in DNA methylation. Selection on the CMJR criterion altered the transcriptomic response of the duodenum after a meal, and resulted in differences in DNA methylation between these two lines.

Table des matières

Remerciements

En premier lieu, je tiens à remercier mon encadrant de stage, *Guillaume DEVAILLY,* pour son offre de stage, son accueil, son soutien et les connaissances qu'il a sues partager avec moi. Je le remercie aussi pour sa disponibilité et la qualité de son encadrement

Je tiens également à adresser mes remerciements à *Julie DEMARS*, représentante de l'équipe GenEpi, pour son accueil chaleureux et sa gentillesse.

J'adresse un grand merci à toute l'équipe GenEpi et à tous les membres de l'unité GenPhySE que j'ai pu voir et à tous ceux que je n'ai pas eus la chance de rencontrer.

Enfin, je désire remercier mes enseignants, *Thomas CLOUAIRE* qui a accepté d'évaluer mon travail et ma famille, notamment *Yacine*, maman et mes sœurs pour leur soutien moral dans ces moments difficiles.

Table des abréviations

CMJR-	Consommation moyenne journalière résiduelle efficace
CMJR+	Consommation moyenne journalière résiduelle peu efficace
G11-	Lignée CMJR- de 11ème génération
G11+	Lignée CMJR+ de 11ème génération
DE	Différentiellement exprimé
TF	Facteurs de transcription
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
CPG	Cytosine-phosphate-Guanine : enchainement d'une Cytosine et d'une Guanine
CNV	Copy number variation
PP	Polypeptide pancréatique
PYY	Peptide YY
CCK	Cholécystokinine
GLP_1	Glucagon-like peptide-1
GIP	Gastric inhibitory peptide
ITPR2	Inositol 1,4,5-trisphosphate receptor type 2
ITGA5	Integrin alpha-5
CFAP97	Cilia and flagella associated protein 97
WDR74	WD repeat domain 74

Introduction

L'entreprise et son secteur d'activité

INRAE, l'institut national de recherche pour l'agriculture, l'alimentation et l'environnement est né le 1er janvier 2020 à l'issu de la fusion entre l'INRA, Institut national de la recherche agronomique et IRSTEA, Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture.

INRAE rassemble une communauté de 12 000 personnes, avec 267 unités de recherche, service et expérimentales implantées dans 18 centres de recherche sur toute la France. L'institut se positionne parmi les tout premiers organismes de recherche au monde en sciences agricoles et alimentaires, en sciences du végétal et de l'animal, et en écologie-environnement. Il est le premier organisme de recherche mondial spécialisé sur l'ensemble « agriculture-alimentation-environnement »

Le centre de recherche INRAE Occitanie-Toulouse compte plus de 978 agents INRAE dont 673 agents titulaires. À l'heure actuelle, il est réparti en 20 structures de recherche dont 11 unités mixtes et 2 unités expérimentales sur 10 implantations géographiques. Il représente plus de 11% des publications de l'Institut.

Les équipes du centre INRAE Occitanie-Toulouse organisent leurs activités sur 4 grands axes :

- La biologie intégrative et prédictive (végétale, animale et micro-organismes)
- Toxicologie alimentaire et santé (animale et humaine)
- Biotechnologie et bio-économie
- Agroécologie des territoires agricoles et forestiers, économie de l'environnement, analyse des filières

L'unité de recherche GenPhySE

L'unité GenPhySE fait partie des départements « génétique animale » et « physiologie animale et systèmes d'élevage » d'INRAE.

L'unité de recherche travaille dans le cadre de la compréhension de la génétique et de la physiologie des animaux d'élevage en prenant en compte leurs environnements et leurs systèmes de production.

Leurs principales activités ont pour but de :

- Explorer la variabilité génétique de caractères complexes chez le bétail
- Augmenter le gain génétique grâce à la sélection génomique et à la conception de nouveaux programmes de sélection
- Améliorer la compréhension des effets environnementaux sur les phénotypes
- Concevoir des systèmes de production animale plus durables

Durant mon projet, j'ai utilisé le cluster de calcul **Genotoul** pour effectuer les tâches et calculs qui nécessitaient beaucoup d'espace mémoire et du temps de calcul. Genotoul est un réseau toulousain de plateformes technologiques et de recherche en sciences du vivant (biologie fondamentale, agronomie, environnement, santé). Les plateformes offrent des compétences et des ressources technologiques de haut niveau. Cet ensemble permet de proposer des études à l'échelle de l'atome jusqu'à celle de la population, en passant par la molécule, la cellule, le tissu ou organe et l'organisme entier.

Contexte et objectif

La filière porcine :

En 2020, près de 4 millions de tonnes de viande porcine ont été consommées par les français(FranceAgriMer, 2021), cette consommation qui ne cesse d'augmenter se traduit par une forte demande, satisfaite en grande partie par une production intensive associée à d'importants impacts environnementaux.

En Europe, les élevages consomment annuellement près de 220 millions de tonnes de céréales et d'oléo-protéagineux dont la moitié sous forme de concentrés industriels riches en protéines et en énergie (Joly, 2021). Cela nécessite des dizaines de millions d'hectares de terres pour subvenir à l'alimentation des cheptels. Cette filière consomme environ 45% de l'énergie utilisée en agriculture et contribue aux émissions de gaz à effet de serre ce qui impacte le changement climatique. L'excès du fumier riche en azote et en phosphate dans les zones denses en élevage entraîne un déséquilibre de l'écosystème et la détérioration de la qualité de l'eau dûe à l'eutrophisation des eaux de ces régions (Dumont et al., 2016). C'est le cas de la Bretagne, première région productrice nationale de porc, qui voit ces cours d'eaux contaminés par du lisier émanant des élevages porcins en causant des dégâts environnementaux sur plusieurs kilomètres. Actuellement, un modèle d'élevage plus soucieux du bien-être animal et de la protection des ressources en eau se voit développer par certains éleveurs : l'élevage des cochons sur de la paille, car ce dernier résorbe une partie de l'azote.

La filière porcine représente 1/3 de l'élevage en France, elle est concentrée sur l'Ouest avec la Bretagne qui représente près de 60% de la production, elle comptait en 2019 un cheptel de 7.6 millions de têtes (Direction Régionale de l'Alimentation, 2020). La France prend la 7ème place dans la production de viande porcine en 2020, soit 2.6% de la production mondiale (3trois3.com, 2020) en précisant qu'elle était autosuffisante à 72% en jambon en 2018 (Le média de l'alimentaire, 2018).
Le coût de l'alimentation est la composante principale (deux tiers ou plus) du coût total de la production de viande de porc(Gondret et al., 2017).

L'efficacité alimentaire

Améliorer l'efficacité alimentaire est devenu un des enjeux majeurs pour contourner ces coûts, réduire l'impact environnemental et répondre aux besoins nutritionnels de la population humaine (Soleimani et al., 2021).

La Consommation moyenne journalière résiduelle (CMJR) est une des mesures d'évaluation de l'efficacité alimentaire. Elle résulte de la différence entre la consommation théorique et la consommation observée pour une production optimale. INRAE a développé sur 11 générations de sélection, deux lignées de porcs divergentes issues d'une même population de race Large White, ces lignées ont été sélectionnées sur le critère de l'efficacité alimentaire, des CMJR- qui avec la même quantité de nourriture consommée, produisent une quantité de viande plus importante que les CMJR+, avec des taux de croissance équivalentes (Soleimani et al., 2021).

De nombreuses études ont été réalisées sur les lignées de porcs CMJR+/CMJR- (Gondret et al., 2017; Gilbert et al., 2019; Soleimani et al., 2021), il a été notamment concerté une différence de comportement entre ces deux lignées.

Les différences de prise alimentaire entre ces deux lignées devraient résulter d'une modification de la régulation de l'appétit et de la satiété, ainsi du fait que les porcs efficaces ont une meilleure capacité à digérer et à absorber les nutriments (gènes de transporteurs de nutriments ou d'enzyme digestives).

La régulation de l'appétit et de la satiété

L'appétit, selon (Blundell et al., 2010) couvre tout le domaine de la prise alimentaire, de la sélection, de la motivation et des préférences. Cette prise alimentaire assure la consommation périodique de substances sources d'énergie et de nutriments tirés de l'environnement. Elle participe de façon essentielle à plusieurs mécanismes homéostatiques (maintien de la glycémie ; régulation du bilan d'énergie) qui réalisent la stabilité du milieu intérieur assurant à l'animal (ou à l'homme) une vie autonome (Bellisle, 2005)

On peut définir aussi l'appétit comme étant le désir de manger une substance particulière qui, dans l'expérience alimentaire antérieure du mangeur, s'est avérée capable de corriger une carence spécifique, en vitamines par exemple (Bellisle, 2005)

La consommation de nourriture entraine l'activation de signaux de satiété, ces signaux remontent de l'intestin vers les structures cérébrales où il va y avoir un rétrocontrôle négatif, ce qui inhibe la satiété. Les signaux augmentent au fur et à mesure que l'individu s'alimente

jusqu'à atteindre le maximum : la prise alimentaire est alors inhibée pour une certaine durée, les signaux vont ensuite disparaitre au fur et à mesure et un nouveau cycle apparait. La satiété est l'état final qui se produit à la fin du repas pour empêcher tout comportement alimentaire ultérieur(Halford and Harrold, 2012).

La satiété a été définie par (Magnen, 2012) comme l'état d'absence de faim, d'absence de désir de manger, s'accompagne généralement d'un état de détente associé à la satisfaction du besoin métabolique. La satiété serait également un état psychophysiologique complexe qui évolue après un repas selon la « cascade de la satiété » proposée originellement par (Blundell et al, 1987). L'intensité et la durée de cet état de satiété dépendent de plusieurs facteurs, dont le contenu énergétique et nutritionnel du repas précédent. (Bellisle, 2005)

La sensation de l'appétit et de la satiété est régulée par la sécrétion des hormones qui varient en fonction de notre prise alimentaire, appelé « hormones de régulation de l'appétit », qui sont classées en hormones orexigènes lorsqu'elles s'associent à une sensation de faim, et anorexigènes pour la sensation de satiété. Si l'on ressent la faim, c'est l'effet de la ghréline sécrétée par l'estomac. Si le repas s'arrête, c'est sous l'effet d'une multitude d'hormones sécrétées par l'intestin, comme le polypeptide YY (PYY, ou peptide tyrosine tyrosine), le glucagon-like peptide-1 (GLP-1), GIP (gastric inhibitory peptide). Ces hormones du tractus digestif agissent dans une boucle de rétroaction qui associe l'hypothalamus et le nerf vague(Galusca et al., 2016). D'autres hormones d'origine pancréatique ou digestive jouent un rôle important dans la régulation de l'appétit telles que : le polypeptide pancréatique (PP), la cholécystokinine (CCK), la leptine, à quoi s'ajoute la sérotonine d'origine digestive et les neuromédiateurs centraux qui intègrent ces signaux.

La méthylation de l'ADN

La méthylation de l'ADN est un des supports de l'épigénétique, impliquée dans plusieurs processus comme la régulation de l'expression des gènes et la formation de la chromatine. Elle constitue un élément clé de la régulation épigénétique de l'expression des gènes (Vilain, 2016). Ce phénomène correspond à l'ajout d'un groupement méthyle en position 5 d'une cytosine, qui la plupart du temps est incluse dans un dinucléotide CpG : c'est une modification chimique de l'ADN qui ne modifie pas la séquence d'acides nucléiques (Cartron et al., 2015).

Chez les vertébrés, ces sites méthylables suivent une distribution non uniforme : il existe des domaines, appelés îlots CpG, où ce dinucléotide est plus fortement représenté (Filion and Defossez, 2004). Ces ilots CpG jouent un rôle important dans la régulation de la transcription génique qui sont généralement situés au niveau des sites d'initiation à la transcription.

Mon projet de stage consiste à étudier et analyser des données de séquençage haut débit concernant la méthylation et la transcription de l'ADN chez deux lignées de porcs sélectionnés et qui divergent sur un critère d'efficacité alimentaire.

Le transcriptome des animaux CMJR+/- a déjà été étudié par (Gondret et al., 2017), ils ont analysé les données transcriptomiques générées à partir du muscle, du foie et de deux tissus adipeux chez deux lignées de porcs sélectionnées sur 8 générations : une lignée à un faible CMJR et l'autre à un CMJR élevé (moins efficace). Leurs principaux résultats étaient que : *(i)* l'effet le plus apparent de la lignée a été observé dans le muscle alors que le tissu adipeux sous-cutané était le tissu le moins impacté, *(ii)* de nombreux gènes immunitaires étaient sous-exprimés dans les quatre tissus des porcs les plus efficaces, *(iii)* les lignées impliqués dans l'oxydation du gras saturé ont été sous exprimés dans le muscle mais surexprimés dans les tissus adipeux de la lignée efficace par rapport à la lignées moins efficace (Gondret et al., 2017). Notre étude s'intéresse à un organe qui n'est pas analysé par cette étude (le duodénum), notre choix s'est porté sur cet organe pour : *(i)* sa position proximale dans l'intestin (juste après l'estomac), *(ii)* son rôle direct dans la régulation de la satiété (sécrétion d'hormones) et *(iii)* son rôle dans la digestion et l'absorption des nutriments.

Afin de mieux caractériser les différences entre les deux lignées sélectionnées, la lignée efficace (CMJR-) et la lignée peu efficace (CMJR+), nous avons comparé le transcriptome et le méthylome du duodénum de porcs issus de ces deux lignées sous une condition alimentaire : à jeun ou après un repas, et ce, dans le but d'identifier les gènes différentiellement exprimés dans le duodénum avant et après la prise alimentaire, et comparer la réponse des deux lignées divergentes. Nous avons également étudié dans quelle mesure le méthylome des deux lignées permettrait d'expliquer d'éventuelles différences de transcriptome.

Notre étude s'est basé sur 24 échantillons issus de 24 porcelets de race Large White en début d'engraissement élevé par les équipes d'élevage d'INRAE, issus de la génération 11 des lignées CMJR (désignées G11+ et G11-). Le dispositif expérimental est plus détaillé dans la partie

Matériels et méthodes. L'expérimentation animale et la préparation des banques de séquençage ont été réalisés avant mon arrivée dans l'équipe.

Matériels et méthodes

Dispositif expérimental

Dans cette partie, nous détaillons l'ensemble du jeu de données sur lequel notre étude s'est basée. Le jeu de données est composé de 24 porcelets en début d'engraissement avec un âge moyen de 60 jours à la date de l'abattage et un poids entre 30 et 35kg. Ils sont issus de 6 portées de parents différents avec 4 porcelets pour chaque portée.

Comme la figure1 nous le montre, nous disposons de 2 un groupe de 12 porcelets avec une forte/faible CMJR.

Le groupe CMJR- est constitué de :

- 3 femelles et 3 males nourris
- 3 femelles et 3 males à jeun

- Le groupe CMJR+ est constitué de :
  - 2 femelles et 4 males à jeun
  - 3 femelles et 3 males à jeun

Ce léger déséquilibre dans l'une des portés est dû au fait que l'une des femelles était malade le matin de l'expérience et a été remplacée par son frère.

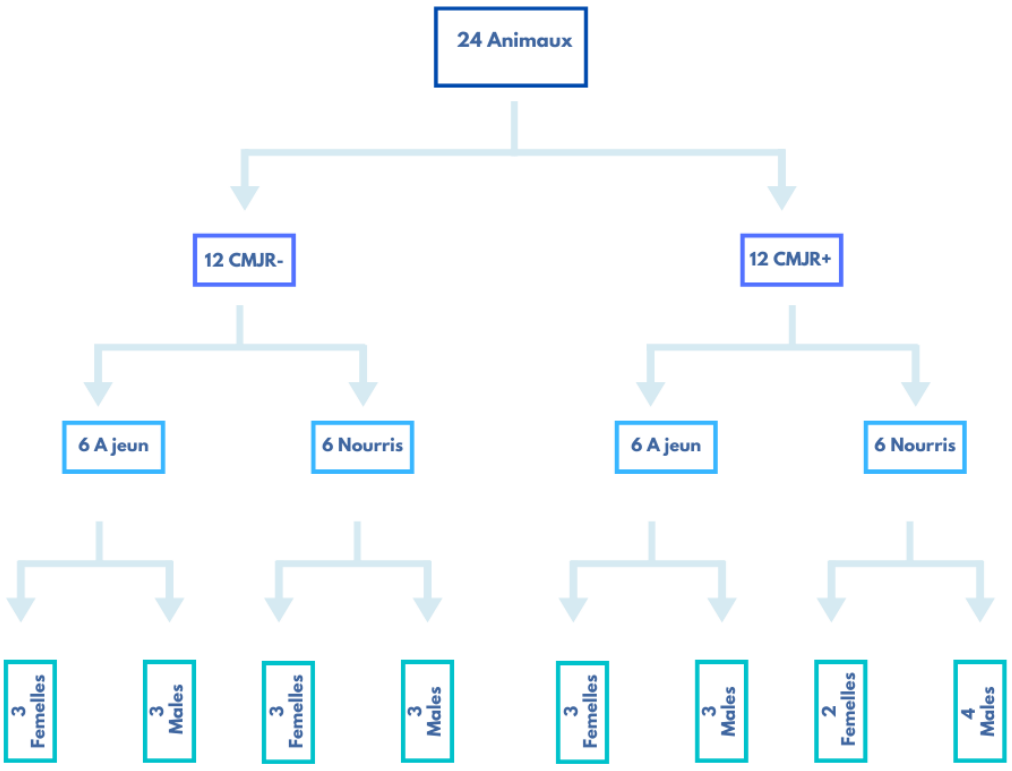

Figure 1 Présentation des échantillons du dispositif expérimental utilisé pour l'étude.

Dans les deux parties qui vont suivre, nous allons utiliser les termes « G11+ et G11- » pour désigner les lignées « CMJR+ et CMJR- » respectivement, et les terme « fasted et fed » pour désigner les conditions « à jeun et nourris » respectivement.

Expérimentations animales

La veille de l'échantillonnage, la nourriture a été retirée des cases des animaux à 16h30. Dans l'une des deux cases, la nourriture a été réintroduite à 8h00 le matin (groupe 'Nourris'). Les animaux avaient accès à volonté à de l'eau de boisson, et les cases disposaient d'un milieu enrichi. L'expérimentation animale a reçu l'agrément ministériel numéro APAFIS#21107-2018120415595562 v10. Suite à l'abattage, une section de 5 cm de longueur du duodénum a été prélevée par dissection, sectionnée en longueur, et lavée abondamment au PBS. La muqueuse duodénale a été récupérée par grattage avec une lame de verre, sectionnée, et congelé dans l'azote liquide.

Préparation des banques de séquençage

Les échantillons de duodénum ont été broyés. La poudre a permis l'extraction d'ARN et d'ADN via les kits « NucleoSpin TriPrep, Mini kit pour ARN, ADN, et purification de protéine » de Macherey-Nagel. Des librairies de séquençages Illumina ont ensuite été préparées par la plateforme Genotoul, à la fois pour le RNA-seq et le MeDP-seq. Les librairies ARN ont été séquencées tels qu'elles. Une précipitation d'ADN méthylé a été préparée sur un mélange des librairies ADN en utilisant le kit *MethylMiner* de ThermoFisher selon les instructions du fabricant. En complément des 24 échantillons, 4 fractions Inputs (ADN non précipités) et 4 fractions artificiellement méthylés via la méthyl transférase M.Sss1 (NEB) ont été préparées pour servir de contrôles négatifs et positifs de la précipitation d'ADN méthylé. Les données de séquençage seront prochainement disponibles sur la base de donnée publique *European Nucleotide Archive*. Le schéma suivant illustre la méthode réalisée.

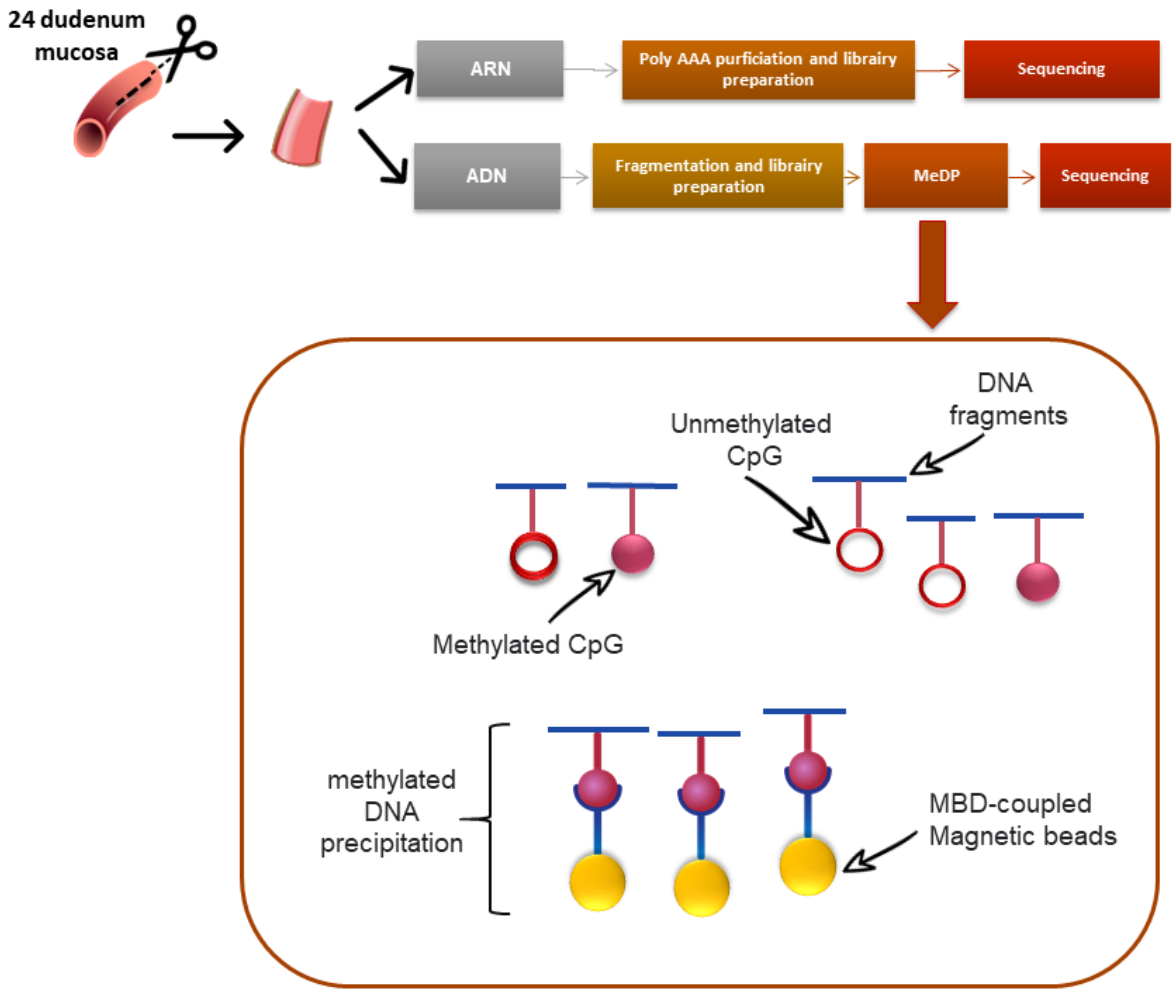

Schéma de la préparation des banques de séquençages en partant des échantillons du duodénum des 24 porcelets.

Positionnement et quantification des lectures

Nf-core RNA-seq

Les pipelines nf-core rassemblent un ensemble organisé de pipelines d'analyse des meilleures pratiques crées à l'aide de Nextflow (Ewels et al., 2020). Ils permettent d'exécuter des tâches sur plusieurs infrastructures de calcul de manière très portable.

Avec les données de séquençage d'ARN obtenues après l'étape précédente, nous avons appliqué le pipeline nf-core/rna-seq en utilisant :

- Une feuille d'échantillons avec des informations sur les échantillons que nous avons souhaité analyser sous forme d'un fichier csv avec 4 colonnes séparés par des virgules comme suit (sample, fastq_1, fastq_2, strandedness) avec :
    - sample : nom de l'échantillon

  - fastq_1 & fastq_2 : chemin d'accès complet au fichier FastQ pour les lectures
  - Strandedness : cpécificité du brin de l'échantillon. Dans notre cas, nous avons utilisé « reverse ».
- Le génome et l'annotation de référence de notre organisme Sscrofa11,1 (téléchargés depuis Ensembl v102 le 12/01/2021).

Nous avons quantifié l'expression des gènes en utilisant le mode pseudo-alignement de Salmon (Patro et al., 2017), et généré des fichiers Bam par alignement STAR afin de visualiser nos résultats sur un navigateur de génome. Les tableaux de comptages par transcrits ont été agglomérés par gènes en utilisant le package Bioconductor **tximport** (Soneson et al., 2015).

Nous avons lancé le pipeline sous forme de job sur le cluster de calcul Genotoul, et après 5h d'exécution, le pipeline nous avait généré un répertoire résultats contenant le fichier de contrôle de qualité (MultiQC) et tous les résultats des tâches demandés.

Nf-core ChIP-seq

Comme pour les données RNA-seq, nous avons procédé de la même manière pour les données de méthylation, cette fois ci, nous avons appliqué le pipeline ChIP-seq pour nos échantillons immunoprécipités. Nf-core ne disposant pas d'un pipeline spécifique à MeDP-seq, nous avons opté pour le pipeline ChIP-seq car les deux types de données sont de nature très similaire.

Comme pour les données de transcription, nous avons besoin :

- D'une feuille d'échantillon contenant les informations sur nos données de méthylations sus forme d'un csv de 6 colonnes séparés par des virgules comme suit (group, replicate, fastq_1, fastq_2, antibody, control) avec :
  - group : identifiant pour l'échantillon
  - replicate : entier représentant le nombre de répliques
  - fastq_1 & fastq_2 : chemin d'accès complet au fichier FastQ pour les lectures
  - antibody : nom de l'anticorps pour séparer l'analyse en aval des différents anticorps, nous avons utilisé l'anticorps « MDB »
  - control : identifiant pour l'échantillon control, « Tem-input » dans notre cas.
- Le génome et l'annotation de référence de notre organisme (le même utilisé pour l'analyse RNA-seq).

En plus des données d'entrées, nous avons mis le paramètre **macs_gsize** à 2.25e9 qui représente la taille effective de notre génome, ce paramètre est requis par MACS2 (un outil couramment

utilisé pour identifier les sites des liaisons des facteurs de transcriptions). Et le paramètre **min_reps_consensus** à 3, cette valeur signifie le nombre de répétitions biologiques requises à partir d'une condition donnée pour qu'un pic contribue à un pic de consensus, dans notre cas, tous les pics qui sont uniques à un échantillon ou 2 seront ignorés. Comme avant, nous avons lancé le pipeline sous forme de job sur le cluster de calcul Genotoul, l'exécution a durée 16h.

Analyse différentielle des données RNA-seq

En premier lieu, nous avons réalisé une analyse différentielle sur les données transcriptomique. On a cherché à voir l'expression différentielle dans nos échantillons selon trois facteurs :

- La lignée : **G11-** vs **C11+**.
- La condition alimentaire : **fasted** vs **fed**.
- Le sexe : **male** vs **femelle**.

Pour ce faire, à partir de nos méta-données et du génome de référence, nous avons construit un modèle linaire en appliquant **Limma-Voom** (Ritchie et al., 2015) **:**

$$expr \sim line + condition + sex$$

Le package Limma a été développé à l'origine pour l'analyse d'expression différentielle des données de puces à ADN. Voom est une fonction du package limma qui modifie les données RNA-seq pour une utilisation avec limma. Ils permettent des analyses rapides, flexibles et puissantes des données RNA-seq (Ritchie et al., 2015).

Le modèle crée un tableau de variables fictives qu'il utilise pour la modélisation statistique. Dans notre cas, il attribue une valeur de logFC (une valeur seuil pour déterminer les gènes différentiellement exprimés) de sorte que :

- Avec le facteur Lignée : une **logFC négative** et une **pval_adjusted < 0.01** signifie une surexpression chez les G11- (DOWN) et une valeur de **logFC positive** et une pval**_adjusted < 0.01** signifie une surexpression chez les G11+ (UP). Les pValues ont été ajustée avec la méthode FDR (False Discovery Rate).
- Avec le facteur Condition : une **logFC négative** et une **pval_adjusted < 0.01** signifie une surexpression avant la prise alimentaire (DOWN) et une valeur de **logFC positive** et une **pval_adjusted < 0.01** signifie une surexpression après la prise du repas (UP).
- Avec le facteur sexe : une **logFC négative** et une **pval_adjusted < 0.01** signifie une surexpression chez les femelles (DOWN) et une valeur de **logFC positive** et une **pval_adjusted < 0.01** signifie une surexpression chez les males (UP).

Nous avons étudié les effets présents sur l'expression de nos gènes en incluant les 3 facteurs. Nous avons aussi cherché d'éventuelles interactions lignée x conditions dans notre modèle de détection de gènes différentiellement exprimés en utilisant le modèle :

$$expr \sim line + condition + line:condition.$$

Annotation des gènes différentiellement exprimés

Avec les gènes différentiellement exprimés, il était intéressant d'avoir une idée sur leurs fonctions et les processus biologiques dans lesquels ils interviennent.

Une des bases de données qui nous a permis de répondre à notre question est PantherDB. Un des rôles principaux de PantherDB est de déduire la fonction des gènes et des protéines sur de grandes bases de données de séquences (Mi et al., 2013).

Pour notre étude, nous avons extrait depuis l'analyse différentielle précédente, les listes des gènes surexprimé chez les G11-, les G11+ pour le facteur Lignée, et les gènes surexprimés avant la prise alimentaire et ceux après pour le facteur Condition. La liste des gènes analysés a servi comme jeu de données de référence. Nous avons chargé ces listes de gènes dans PantherDB, choisi l'organisme de référence et sélectionné l'analyse souhaitée (dans notre cas on s'intéresse plus aux processus biologiques dans les quels nos gènes apparaissent, même si on a pu regarder les autres domaines pour satisfaire notre curiosité). Nous avons regardé les catégories GO enrichies parmi les "Uniquely Mapped ID".

Dans le but d'identifier les facteurs de transcriptions (TFs) critiques à partir de nos données d'expression géniques, nous avons opté par une métrique modélisée par (Reverter et al., 2010) nommé RIF (Regulatory Impact Factor). RIF est une métrique donnée à chaque TF qui combine le changement de co-expression entre les gènes TFs et DE, elle est représentée avec deux mesures :

- RIF1 : il est attribué aux TFs qui sont systématiquement le plus différentiellement co-exprimés avec les gènes hautement abondants et hautement DE.
- RIF2 : il est attribué aux TFs dont la capacité à prédire l'abondance des gènes DE (RIF1) est la plus altérée.

Après les premiers TFs détectés avec RIF, Nous voulions nous assurer de la puissance de RIF, de voir qui nous pouvions retrouver et confirmer les TFs précédemment détectés. Nous avons procédé avec une autre approche différente qui se base sur les positionnements des motifs des

gènes DE, nous avons utilisé ***g:profiler*** , un serveur web qui permet de caractériser et manipuler des listes de gènes avec une interface graphique simple et permettant une visualisation puissante (Raudvere et al., 2019). Nous avons utilisé l'outil ***g:orth*** afin de cartographier des gènes orthologues humains basés sur les informations extraites de la base de données Ensemble après avoir extrait les gènes homologues à nos gènes DE.

Régions différentiellement méthylées

La méthylation de l'ADN est un élément essentiel pour la régulation de l'expression génique (Riebler et al., 2014). En théorie, des régions régulatrices hyper méthylées induisent une répression de la transcription et dans le sens inverse, des régions régulatrices hypo méthylées peuvent activer la transcription d'un gène. L'équipe de (Riebler et al., 2014) a développé une approche empirique de bayes qui utilise un échantillon de contrôle entièrement méthylé pour transformer le nombre de lectures observées en proportion de régions méthylées. Cette approche est nommée BayMeth. Nous avons appliqué leur méthode sur nos données afin de déterminer les régions méthylées et leurs proportions en CpG sur nos échantillons d'intérêts (générés par le programme d'alignement du pipeline nf-core/chipseq) en se basant sur un échantillon contrôle entièrement méthylé.

Après l'application de BayMeth, nous avons réussi à utiliser les témoins méthylés pour convertir les profils de MeDP-seq avec des densités de méthylations en profils de taux de méthylation.

Une des analyses classiques des données de comptage de méthylation est de pouvoir détecter les régions différentiellement méthylées. Dans cette optique, nous avons opté pour deux approches assez différentes mais qui sont complémentaire :

- **DMRseq** : une approche basée sur la permutation pour détecter les régions différentiellement méthylées en utilisant des modèles généralisés des moindres carrés (Korthauer et al., 2017), initialement conçu pour des données de séquençage du Bisulfite.
- **DiffBind** : une méthode pour calculer les sites différentiellement méthylés à partir de plusieurs expériences ChIP-seq (Ross-Innes et al., 2012). DiffBind requière une liste de pics en entrée. Pour cette étude, nous avons utilisées trois types d'entrées différentes : la liste de pics MACS2 issus du pipeline nfcore/chipseq, la liste de pics consensus issus du pipeline nfcore/chipseq, contenant moins de pics mais plus larges, ainsi que la liste

des îlots CpG porcin issue de la « UCSC table browser » (Karolchik et al., 2004) qui ont une forte probabilité d'êtres des régions régulatrices.

Pour chacune de ces méthodes, nous avons appliqué une greyListe afin d'avoir une analyse plus propre, l'idée de notre greyListe était de filtrer et d'éliminer les CNV (copy Number Variation) et autres régions sur-séquencées (Brown, 2021). Le package Bioconductor grayListChip utilise les témoins inputs pour identifier les régions sur-séquencées susceptibles d'être des CNV.

Disponibilité du code, des données et développement de package

Nous avons tout au long du projet, tenu un répertoire **GitLab** nommé ***rosePigs*** forgemia.inra.fr/safia.saci/rosepigs contenant nos scripts, et les résultats de nos analyses avec une documentation qui nous permettait un échange facile et rapide.

Nous avons également pu développer deux packages R dont un d'annotation de données (génome Ensembl porcin, que j'ai réalisé car seul la version UCSC est disponible sur Bioconductor) et l'autre de visualisation nommé **Epistack**, ce dernier, toujours en cours de développement, sera soumis à Bioconductor github.com/GenEpi-GenPhySE/epistack

Epistack est un package permettant la génération d'une visualisation informative très utilisée en bio-informatique : les piles de profils épigénétiques centrées sur un type de région donnée (promoteurs de gènes, sommets de piques, etc.). Il se veut (un peu) plus simple d'utilisation que les méthodes alternatives existantes, tout en restant suffisamment flexible pour s'adapter à un grand nombre de situations. Une attention particulière a été portée au problème d'*overplotting* lorsqu'un grand nombre de régions (> 10 000) est visualisée en même temps. Nous avons utilisé Epistack dans notre étude pour générer automatiquement un grand nombre de figures pour visualiser nos données MeDP-seq, ainsi que pour générer les figures 5B et 7 de ce rapport.

Résultats

La réponse transcriptomique à la prise alimentaire est altérée par la sélection sur l'efficacité alimentaire :

La première approche a été d'étudier et analyser le transcriptome du duodénum de porcs avant et après un repas chez deux lignées divergeant sur un critère d'efficacité alimentaire.

Nous avons analysé en premier lieux les données RNA-Seq afin de déterminer les gènes différentiellement exprimés (DE) en fonction de 3 facteurs : la lignée (G11- vs G11+), la condition (à jeun vs nourris) et le sexe (male vs femelle). Nous avons constaté 962 gènes DE pour le facteur de la condition (par exemple, *WDR74*, figure 2C, c'est une protéine régulatrice impliqué dans la synthèse de la sous-unité ribosomale 60S), 318 gènes DE pour la lignée (par exemple, *ADCY4*, figure 2C) et seulement 4 gènes pour le facteur sexe (figure2 A et B), ce qui fait que la réponse transcriptomique du duodénum à la prise alimentaire est plus importante que la différence de transcriptome entre les deux lignées de porcs et que le sexe n'a pratiquement pas d'effet. Le choix des gènes donnés en exemple dans la figure 2C ont été sélectionné sur la base de leurs valeur d'expression.

Une analyse par composante principale sur les gènes différentiellement exprimé a été réalisé. Nous constatons que l'axe 1 permet une séparation des traitements (à jeun vs nourris) et l'axe 2 une séparation des lignées G11+/-. La réponse au passage du bol alimentaire semble alors plus forte dans la lignée G11- (haute efficacité alimentaire) que dans la lignée G11+ (figure 2D). Cette différence de réponse transcriptome est aussi visible sur la heatMap (figure 2B).

La listes des gènes différentiellement exprimés sont disponible dans le répertoire gitlab forgemia.inra.fr/safia.saci/rosepigs/-/blob/master/data/rosePigs_genes_table_DE.csv.

Nous avons également pu constater que 16 de ces gènes sont différentiellement exprimés à la foi sous l'effet de la lignée et de la condition (figure 2A), c'est le cas des gènes ITPR2 (un gène qui interviendrait dans la croissance, dans le métabolisme énergétique et dans l'apoptose) et *ITGA5* (figure 2C, connues pour participer à la signalisation médiée par la surface cellulaire).

Nous avons inclus un effet d'interaction lignée x condition dans notre modèle de détection des gènes différentiellement exprimés mais n'avons détecté aucune interaction statistiquement significative après correction des tests multiples. Une des meilleures interactions est montrée en figure 3C(*CFAP97*).

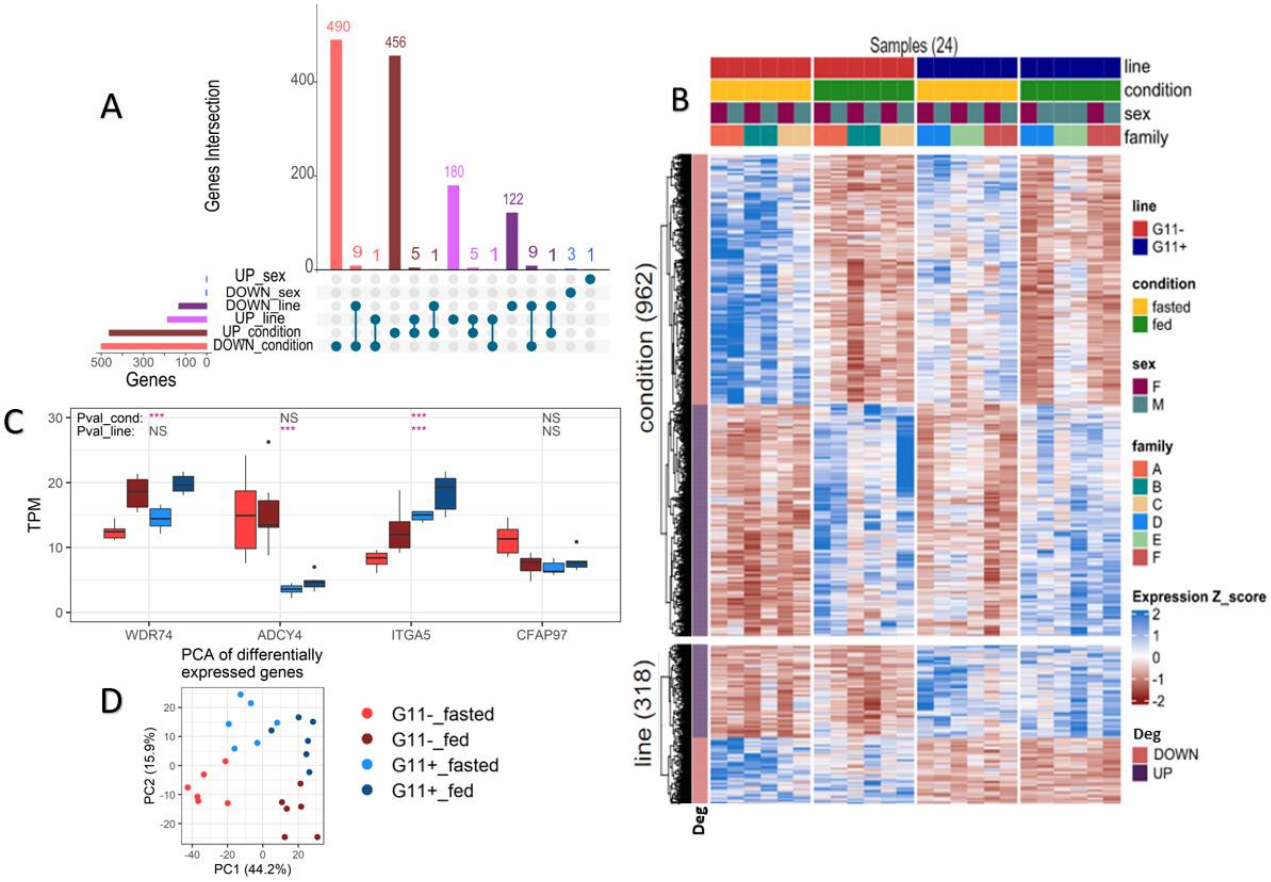

Figure 2 Représentation graphique de l'analyse différentielle : A- Le upSet représente le nombre de gène DE pour chaque facteur sous les deux conditions à chaque fois (représenté par un point bleu unique dans la matrice) et les gènes DE chez deux facteurs simultanément (représenté par deux points bleus reliés par un trait). B- Expressions relatives des gènes différentiellement exprimés, sous forme d'HeatMap annotée. La heatMap nous permet de voir de que chez les G11-, nous avons une réponse plus importante que chez les G11+, également, avec la lignée, nous avons plus de gènes surexprimés avant la prise alimentaire qu'après C- nous avons représenté en boxplots 4 gènes exemples des différences d'expression. D- Analyse par composante principale des gènes différentiellement exprimés.

Annotation des termes enrichis :

Nous avons décidé de cibler nos résultats et voir l'expression des hormones de régulation de l'appétit tel que la *GHRL*, *GLP_1*, *CCK* et *GIP*, L'expression augmente avec la prise de repas chez les G11- mais pas chez les G11+ (figure 9 en annexe).

Comme nous pouvons le voir dans la figure 3, les gènes responsables de la synthèse, sécrétion et inactivation des incrétines (hormones gastro-intestinales qui stimulent la sécrétion d'insuline lorsque la glycémie est trop élevée et qui ralentissent également la vidange gastrique) sont différentiellement exprimés différemment en fonction des CMJR. Par exemple, l'hormone *GCG* (hormone de satiété) est surexprimé avant la prise alimentaire chez la lignée peu efficace et surexprimé après le passage au bol chez la lignée efficace, mais ces changements ne sont pas statistiquement significatifs.

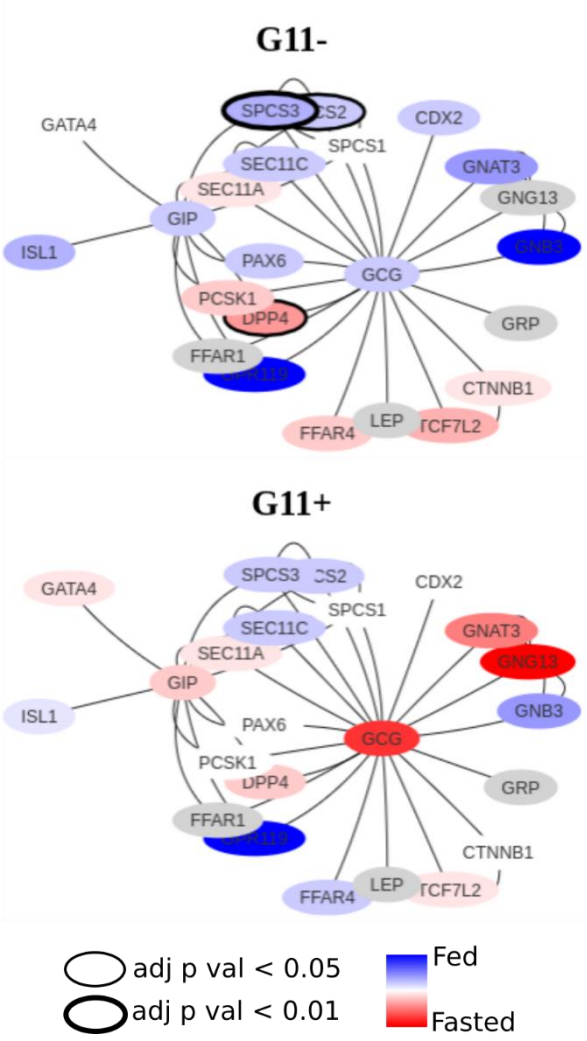

Figure 3 La voix de signalisation "synthèse, sécrétion et inactivation des incrétines" (Reactome R-HSA-400508.1) est représenté sous forme de réseau simplifié. Les nœuds sont colorés selon leur expression moyenne relative avant/après le repas (rouge : expression plus forte avant le repas, bleu : expression plus forte après le repas) dans la lignée G11- (en haut) et G11+ (en bas). Les gènes non détectés par le RNAseq sont en gris. L'épaisseur des ovales reflète la valeure p ajustée (correction FDR) d'un test t sur les TPM des gènes impliqués dans le réseau.

A partir des premiers résultats, nous avant fait une annotation ontologique des gènes DE pour identifier les processus biologiques régulés. On a constaté que la grande majorité de ces fonctions sont enrichies sous le critère de la condition alimentaire et non pas sous le critère de la CMJR (figure 4-B). Les fonctions des gènes surexprimés dans le duodénum à jeun (Condition_down) sont enrichis et impliqués dans le catabolisme (figure 8 en annexe), l'animal mobilisant ses réserves pour compenser l'absence de nourriture. Les fonctions des gènes enrichies dans la condition « nourris » sont eux impliqués dans la synthèse et la maturation d'ARN et de protéines, le métabolisme redémarre après la brève période de jeune.

Aucune ontologie n'a été détectée comme enrichie dans la catégorie down régulée par lignée. Les gènes surexprimés dans la lignée G11+ sont enrichies en gènes impliqués dans la localisation des chromosomes, sans que l'implication biologique de ce résultat ne soit très claire pour l'instant, ce que nous pouvons voir dans la figure 8 en annexe.

Identifications de facteurs de transcriptions régulant les gènes différentiellement exprimés

Les facteurs de transcription (TF) jouent un rôle régulateur central dans la régulation de l'expression des gènes, mais leur détection à partir des données d'expression est limitée en raison de leur expression faible et souvent clairsemée(Reverter et al., 2010). Nous avons utilisé deux approches complémentaires pour identifier des facteurs de transcription impliqués dans la régulation des gènes différentiellement exprimés : une approche basée sur des réseaux de co-expressions (RIF) et une approche basée sur l'enrichissement de motifs au niveau des promoteurs des gènes via *g:Profiler*.

La force de la métrique de facteur d'impact réglementaire (RIF) réside dans sa capacité à intégrer simultanément trois sources d'information en une seule mesure: *(i)* le changement de corrélation existant entre l'expression du TF et les gènes DE; *(ii)* la quantité d'expression différentielle des gènes DE; et *(iii)* l'abondance des gènes DE.

L'analyse RIF permet d'ordonner les facteurs de transcription les plus impliqués dans notre réseau de co-expression. Les meilleurs TF sont représentés en figure 4A. On y constate que ce ne sont pas les même TFs impliqués dans les différences d'expressions par lignée et par conditions, à l'exception de MYBL2, PROX2, PRMT2 ou TEAD1. RIF1 et RIF2 donne des résultats différents.

Nous avons, dans le but de voir si nous pouvions retrouver les TFs identifiés avec RIF, refait une analyse avec Gprofiler. Nous n'avons trouvé aucun TF en commun. Par ailleurs, nous avons constaté que Gprofiler identifie plus de TF impliqués dans la liste de gènes DE par condition que par lignée (figure 4B). L'expression différentielle par lignée impliquerait donc moins d'activité différentielle de TF que l'expression différentielle par condition. Les expressions différentielles par lignée sont sans doute beaucoup plus dues à des différences de séquences génétiques.

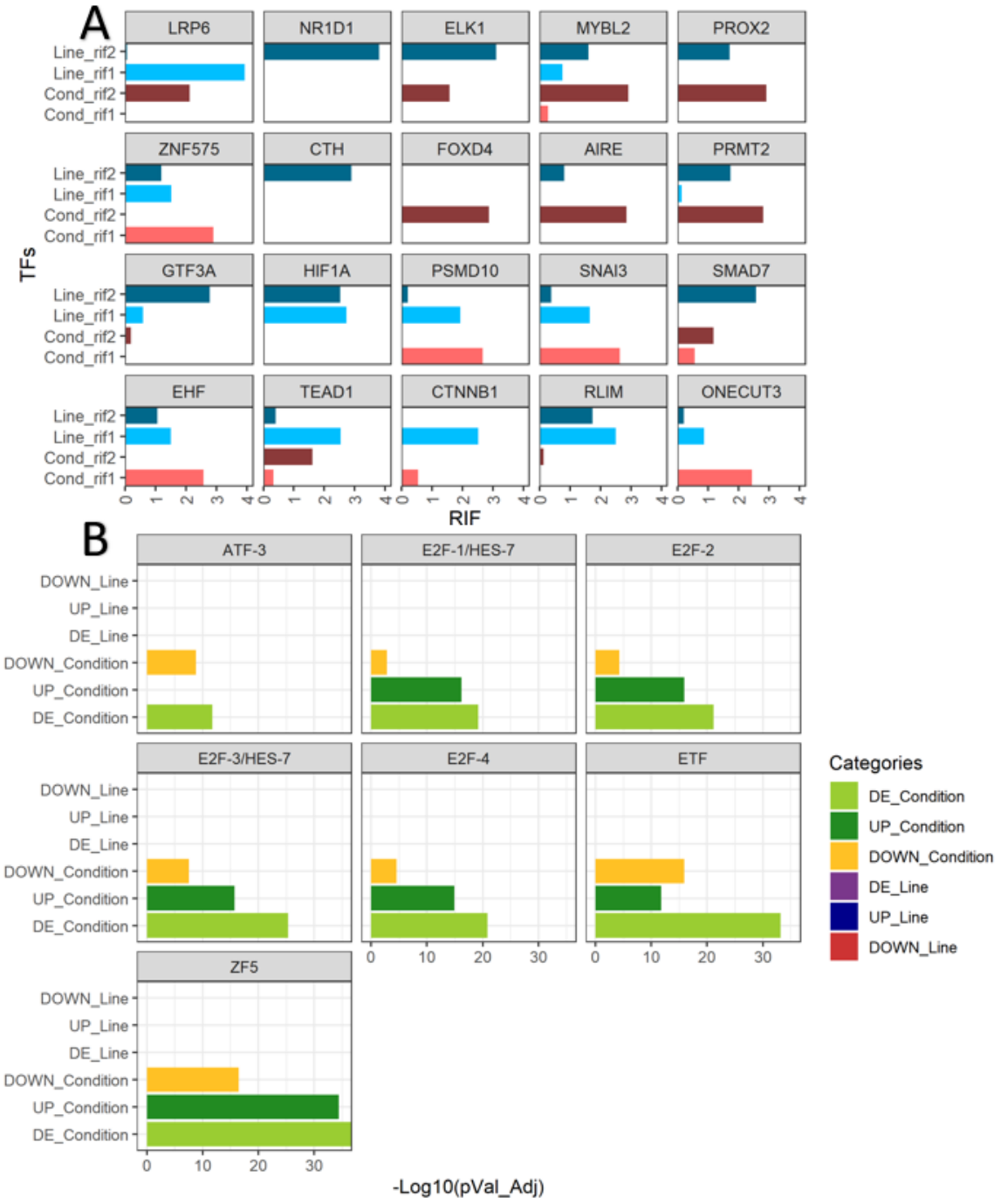

Figure 2 : Représentation des meilleurs facteurs de transcription des gènes différentiellement exprimés : A- Le premier tableau (en haut) représente les premiers TFs renvoyé par l'analyse RIF responsables de l'expression différentielle identifiés et représentés ici avec des bar Plot en se basant sur un zscore. En bleu (claire et foncé), nous avons les meilleurs TFs en fonction des deux score RIF1 et RIF2 pour les gènes DE sur le critère CMJR, en rouge (claire et foncé), nous avons les meilleurs TFs en fonction des scores RIF1 et RIF2 sur les gènes De sur le critère de la condition alimentaire. B- Le deuxième tableau (bas), désigne les meilleurs TFs responsables de l'expression différentielle identifiés avec Gprofiler représentés avec la –log10(Pval_adjusted). En jaune, les TFs responsables de l'expression différentielle avant la prise du repas. En vert foncé, ceux TFs responsables de l'expression différentielle après la prise du repas et en vert claire les TFs issus de la liste des gènes DE sur le critère de la condition alimentaire. Avec g : profiler, nous ne détectons aucun TF sur le critère CMJR.

Identification des régions différentiellement méthylées

La méthylation de l'ADN est impliquée dans la régulation de l'expression des gènes. Nous avons cherché à savoir si les différences transcriptomiques observées avant et après un repas ou entre les deux lignées étaient en partie associées à des différences de méthylation de l'ADN.

Les données de précipitation de l'ADN méthylé nous apportent une information sur la densité de méthylation de l'ADN d'une région donnée. Cette densité de méthylation dépend à la fois de la densité en CpG, variable entre les différentes régions du génome, et le taux de méthylation des sites CpG, variant entre 0 et 1. Nous avons utilisé BayMeth (Riebler et al., 2014), qui permet de reconstruire une information de taux de méthylation à partir de données MeDP-seq et de témoins artificiellement méthylés. L'approche est illustrée sur la figure 5A à partir de 3 régions du génome. On constate que la région '**a'** a une forte couverture en MeDP-seq comparé aux régions '**b'** et '**c'**, et donc une plus forte densité de CpG méthylés. Dans un témoin artificiellement méthylé, les régions 'a' et '**c'** sont plus fortement précipitées que la région '**b'**. On en déduit alors que la région 'b' a un taux de méthylation élevé mais une faible densité en site CpG, tandis que la région '**c'** a un faible taux de méthylation mais une forte densité en sites CpG. Il s'agit d'ailleurs d'un îlot CpG. Nous pouvons voir au niveau de l'ilôt CpG que la méthylation du gène *IGF2BP1* (pris au hasard dans le génome) est à très faible niveau.

Avec nos données meDIP-seq, la densité de méthylation au niveau de promoteur des gènes est négativement corrélée avec l'expression des gènes dans chacun de nos échantillons. Également, après la transformation BayMeth, les résultats étaient aussi concluant avec une forte corrélation entre le niveau de méthylation et l'expression génique (figure 5B).

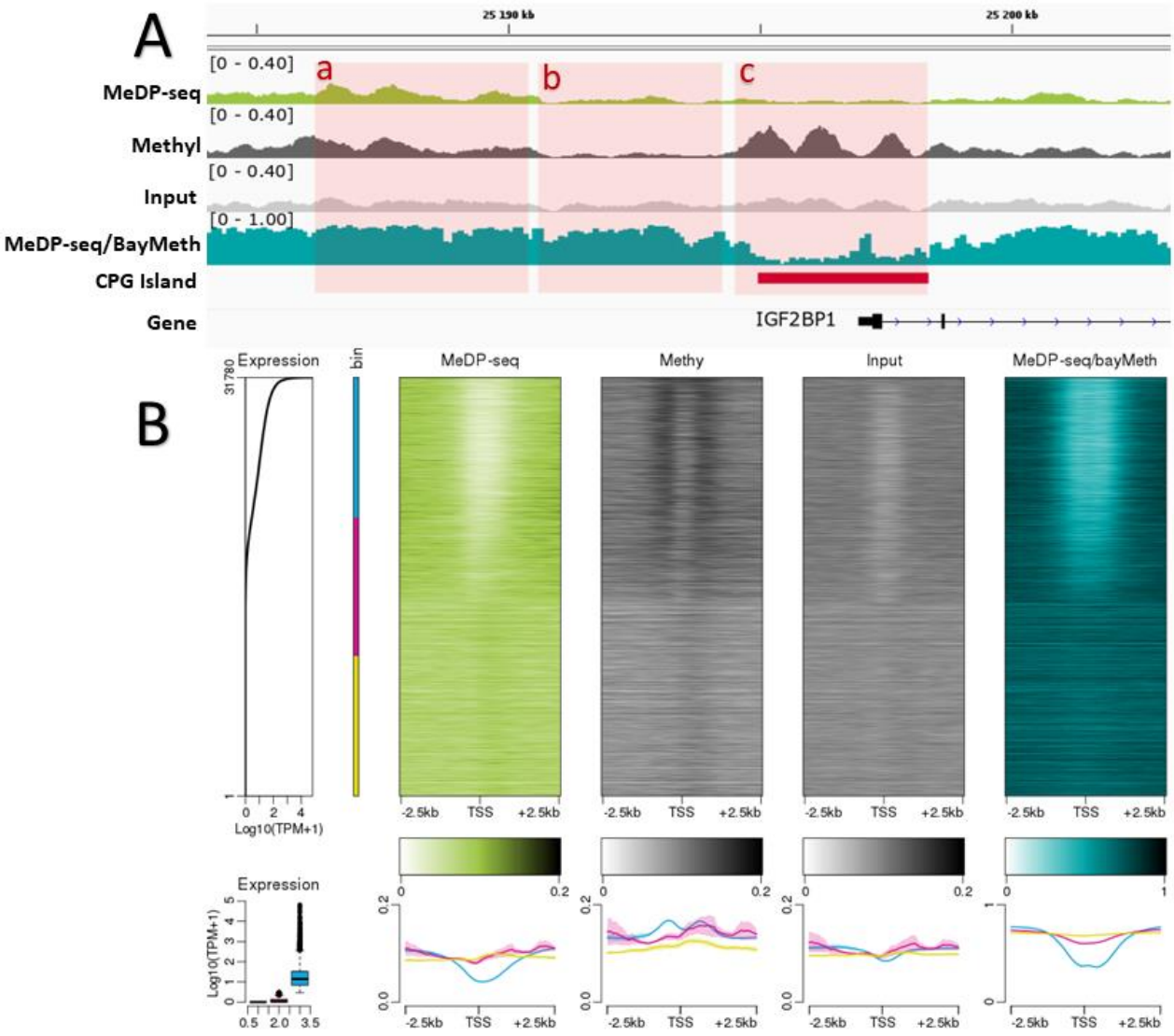

Figure 3 : Représentation des différence de méthylation d'un échantillon efficace après la prise alimentaire aux proximités du gènes IGF2BP1 : A– la piste IGV nous montre la différence des niveau de méthylation de l'échantillon en amont et au centre su site d'initiation à la transcription avant et après la conversion des densité de méthylation en proportion de méthylation. B- Les heatMaps réalisés avec notre package Epistack nous permettent la visualisation des profiles méthylés au niveau des TSS avec l'expression de leurs profiles moyens avant et après la transformation avec BayMeth. La figure représente les données issue d'un proc G11- nourris. Ligne du haut, de gauche à droite : niveau d'expression des gènes d'après le RNA-Seq. Les gènes sont triés du plus exprimé au moins exprimés. Les gènes sont regroupés en trois groupes : en bleu les gènes exprimés, en rouge les gènes peu exprimés, en jaune les gènes non exprimés. Puis: signaux MeDP-seq au niveau des sites d'initiations de la transcription (+/- 2.5kb) pour tous les gènes, triés selon leur niveau d'expression. Puis signaux du témoin artificiellement méthylé, du témoin input, et de la proration de méthylation obtenue par BayMeth. Ligne du bas: Boxplots des valeurs d'expressions pour chacun des trois groupes de gènes, puis profils moyens des signaux MeDP-seq, Témoin méthyl, témoin input et signaux BayMeth.

L'impact de la prise alimentaire sur le méthylome du duodénum

Nous avons procédé à la détection des régions différentiellement méthylées (DMRs) voisines aux ilots CpG avec deux méthodes (DMRseq et DiffBind) (Ross-Innes et al., 2012; Korthauer et al., 2017).

Avec DMRseq, nous avons compté 242267 DMRs en fonction de la lignée dont seulement 514 jugés significatifs (qval < 0.05) et 71831 DMRs en fonction de la condition alimentaire mais aucun n'est significatif. Les résultats de DiffBind donnent plus de DMRs avec la liste des pics consensus (1855 DMRs) et moins avec les listes individuelles des pics (563 DMRs) ce que nous pouvons voir sur la figure 6A. En termes de taille, DMRseq semble détecter des DMRs plus grands que ceux détectés par DiffBind sur les ilots CpG, mais semble plus petits que ceux détectés avec les pics consensus toujours avec l'analyse DiffBind (figure 5B).

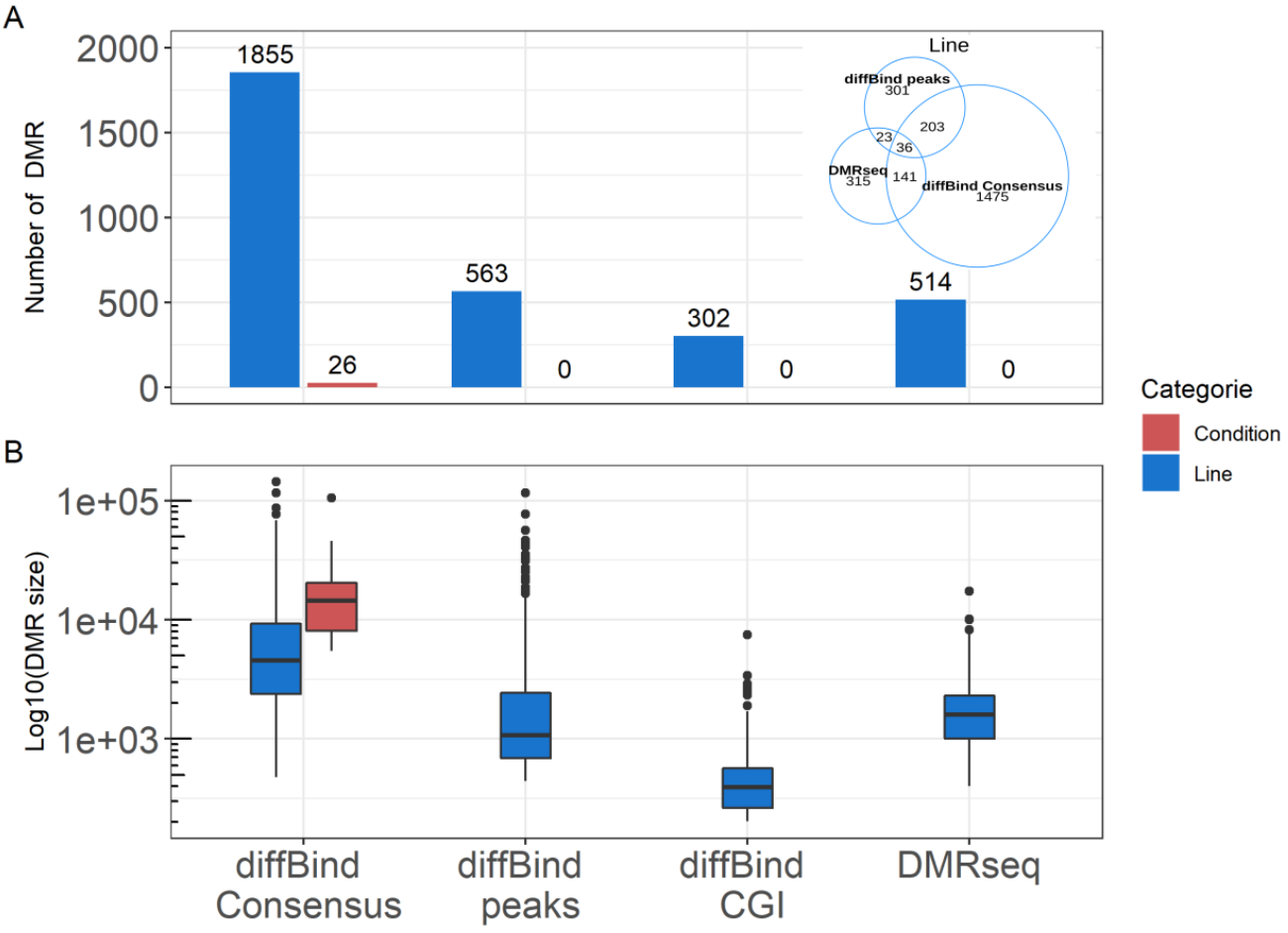

Figure 4 : Représentation des régions différentiellement méthylées avec DMRseq et DiffBind : A- représente le nombre de DMRs significatifs (qval < 0.05) détectés avec DiffBind de gauche à droite sur les pics consensus, sur la liste des pics individuelles et dur la liste des ilots CpG et les DMRs détectés avec DMRseq (bar plot du droite). Sur le diagramme de Venn, nous avons le nombre de DMRs en commun entre les DMRs de DMRseq et ceux de DiffBind. B- représente les tailles en Pb de ces dits DMRs. Les 2 analyses réalisés n'ont pas détecté des DMRs sur le critère de la condition alimentaire.

Le méthylome n'a donc pratiquement pas été impacté par la prise alimentaire, en revanche, la sélection sur 11 génération sur le critère CMJR modifie le méthylome du duodénum de nos lignées (figure 6).

Comme nous pouvons le voir sur le diagramme de Venn de la figure 6-A, nous avons observé que certains DMRs étaient identifiés dans les différentes analyses. Par exemple nous avons 36 DMRs qui sont identifié par DMRseq, par DiffBind avec les broadpeaks en entrée, et même avec l'échantillon consensus. La majorité des DMRs détectées par chaque approche est cependant spécifique de la méthode utilisée.

Nous avons représenté les DMRs identifiées avec DMRseq et DiffBind avec Epistack pour mieux visualiser les profiles des DMRs. Nous avons observé une différence entre les inputs avec DMRseq et dans une moindre mesure avec DiffBind, ce qui reflète la présence de faux positifs parmi nos DMRs, possiblement dû à des CNV ou des artefacts expérimentaux. La figure 7 illustre les DMRs trouvés par DiffBind sur les pics consensus et par DMRseq sur tous le génome.

Pour aller plus loin dans notre analyse, nous avons représenté graphiquement les gènes proches de nos DMRs pour voir si ces derniers avaient un effet sur les gènes voisins. Nous avons constaté que les DMRs avaient des effets contrastés sur l'expression des gènes proches de ces DMRs, ce que nous pouvons voir sur la figure 10 en annexe : beaucoup de gènes proches des DMR ont une expression stable, certains sont surexprimés, d'autres sous exprimés, sans déséquilibre entre les hyperméthylations et les hypométhylations.

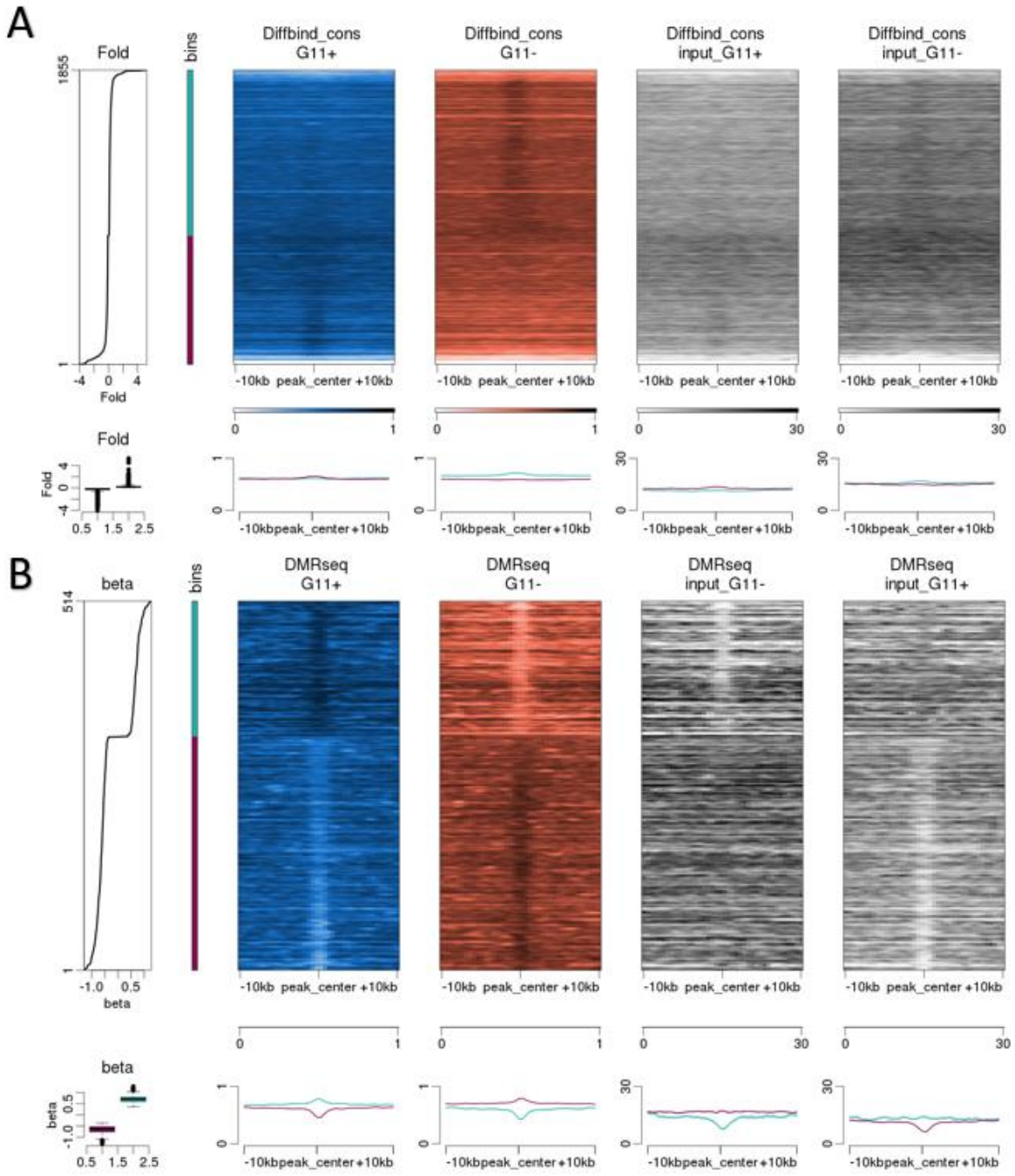

Figure 7 : Les heatMaps ont été réalisés avec Epistack, elles nous permettent la visualisation des profils de méthylation au niveau des DMRs sur les pics consensus détecté avec DiffBind (A) et des DMRs détectés avec DMRseq (B) en comparant les G11+ aux G11-. Les DMRs sont triés avec les hyperméthylations en haut (fold > 0 ou beta > 0, bin : cyan), les hyperméthylations en bas (fold < 0 ou beta < 0, bin : rose foncé). Les profils moyens de méthylation au niveau des centres des DMRs (+/- 10kb) sont représentés pour les G11+ (en bleu), les G11- (en rouge) et pour les témoins inputs des G11+ et des G11- (en gris). Ligne du bas: Boxplots des log2(Fold change) (Fold) ou valeurs betas (beta) pour les hyperméthylations et hyperméthylations, puis profils moyens des signaux de méthylation au niveau des DMRs hypométhylés (rose foncé) et hyperméthylés (cyan).

Discussion

Mon stage avait comme objectif d'étudier le transcriptome et le méthylome avant et après un repas chez deux lignées de porcs qui divergent sur un critère d'efficacité alimentaire. Pour ce faire, nous avons : *(i)* identifié les gènes différentiellement exprimés chez les deux lignées avant et après la prise alimentaire, *(ii)* annoté les termes enrichis des gènes DE pour s'informer des processus biologiques dans lesquels ils interviennent et identifier les facteurs de transcriptions régulant les gènes DE, *(iii)* étudié l'impact de la prise alimentaire sur le méthylome du duodénum de nos lignées sélectionnées.

Nous avons caractérisé le transcriptome du duodénum avant et après la prise alimentaire pour identifier les modifications transcriptomiques dans un organisme qui est plus proche de l'humain que la souris/rat. Ces données et résultats seront précieux pour mieux comprendre la régulation de l'appétit.

Nos résultats nous avaient montré que le transcriptome chez les animaux à forte efficacité alimentaire réagissait plus à la prise alimentaire que chez les animaux à faible efficacité alimentaire. Ce résultat nous semble assez logique vue que les lignées ont été sélectionnées de sorte que les G11- arrive à la sensation de satiété plus vite que les G11+. En revanche nous n'avons pas pu trouver d'interactions lignée x condition avec les données de transcriptions, nous pensons que ceci est dû au fait que nous ne disposons pas d'assez de données donc une faible puissance statistique. En effet, la prise en compte du terme d'interaction dans le modèle statistique **limma** fait passer nos effectifs de 12 échantillons par groupe à seulement 6 par groupe.

La détection des facteurs de transcriptions avec RIF et g:profiler n'ont pas donné des résultats similaires, En effet, RIF se base sur les réseaux des gènes, alors que g:profiler se base sur les motifs des facteurs de transcriptions. Cela ne nous aide pas à aller plus loin dans notre analyse. g:profiler trouve plus de TFs impliqués dans les gènes DE par condition que par lignée. Un résultat pertinent car les gènes DE par lignée ont une source plus génétique alors que les gènes DE par condition ont plus une source de régulation de la transcription via des TFs, nous ne pouvons pas voir ce déséquilibre avec RIF car ce dernier ne fait que classer les TFs et donc nous avons toujours un « top5 » des TFs sans pour autant qu'ils soient les plus pertinents.

Nous avons également conclu que le transcriptome du duodénum de nos animaux a été modifié au passage alimentaire mais pas le méthylome, car l'échantillonnage a été réalisé 12h après le

dernier repas, or, il faut sans doute plus de temps pour modifier le méthylome. La régulation de l'expression des gènes après l'ingestion d'aliments est donc probablement régulée par des facteurs de transcriptions plutôt que par la méthylation de l'ADN.

En identifiant les gènes DE voisins aux DMRs, nous avons conclu qu'il n'y avait pas d'effet entre les DMRs et leurs proches gènes DE. On s'attendait à voir les gènes hyperméthylé sous exprimés et les gènes hypométhylés surexprimés. Nous avons la majorité des gènes qui n'était pas différentiellement exprimée (points gris de la figure 10 en annexe) et quelques-uns qui étaient surexprimés ou sous exprimés (points rouges ou bleus respectivement de la figure 10 en annexe).

Comme perspective à cette étude, nous pouvons aller plus loin dans nos analyses et regarder l'impact des DMR, essayer de comprendre et d'éliminer les faux positifs dans nos DMRs avec d'autres outils comme **diffReps,** un programme développé en PERL et qui fonctionne sur toutes les plates-formes en ligne de commande, conçu pour détecter ces sites différentiels à partir des données ChIP-seq sans partir d'une liste de pics prédéfini (contrairement à DiffBind) , avec ou sans répliques biologiques (Shen et al., 2013). Nous pouvons également essayer de détecter les variations génomiques depuis les données RNAseq/ MeDPseq pour corréler les variations génomiques avec les modifications du transcriptome et du méthylome observés. Nous pouvons également chercher à mieux comprendre les modifications du transcriptome avant et après un repas, cibler les gènes de transporteurs de nutriments ou d'enzyme digestives sur le critère CMJR et pouvoir comparer avec les données rats/souris existantes.

Conclusion

Après avoir analysé le transcriptome et du methylome du duodénum des deux lignées de porcs divergents sur un critère d'efficacité alimentaire, nous pouvons conclure sur 4 points essentiels :

- Le transcriptome du duodénum est modifié par la prise alimentaire.
- La lignées efficace a une réponse transcriptomique au repas plus importante que la lignée moins efficace.
- La méthylation de l'ADN du duodénum n'est pas été impacté par la prise alimentaire.
- Le méthylome du duodénum a été modifié par la sélection divergente par critère CMJR.

Références

3trois3.com. 2020. Statistiques de production porcine - 3trois3, Le site de la filière porc. https://www.3trois3.com/statistiques_porcines/graficos/ (accessed 28 May 2021).

Bellisle, F. 2005. Hunger and satiety, control of food intake. doi: 10.1016/j.emcend.2005.08.003.

Blundell, J., C.D. Graaf, T. Hulshof, S. Jebb, B. Livingstone, et al. 2010. Appetite control: methodological aspects of the evaluation of foods. Obesity Reviews 11(3): 251–270. doi: https://doi.org/10.1111/j.1467-789X.2010.00714.x.

Brown, G. 2021. GreyListChIP: Grey Lists -- Mask Artefact Regions Based on ChIP Inputs. Bioconductor version: Release (3.13).

Cartron, P.-F., R. Pacaud, and G. Salbert. 2015. Méthylation/déméthylation de l'ADN et expression du génome. Revue Francophone des Laboratoires 2015(473): 37–48. doi: 10.1016/S1773-035X(15)30158-1.

Direction Régionale de l'Alimentation, de l'Agriculture et de la F. de B.-S. officiel du service régional du ministère en charge de l'agriculture. 2020. Filière Porcs - Édition 2020. https://draaf.bretagne.agriculture.gouv.fr/Porcs-Edition-2015 (accessed 26 May 2021).

Dumont, B., P. Dupraz, J. Aubin, M. Benoit, V. Chatellier, et al. 2016. Rôles, impacts et services issus des élevages en Europe. Synthèse de l'expertise scientifique collective. auto-saisine.

Filion, G., and P.-A. Defossez. 2004. Les protéines se liant à l'ADN méthylé : interprètes du code épigénétique. Med Sci (Paris) 20(1): 7–8. doi: 10.1051/medsci/20042017.

FranceAgriMer. 2021. PUBLICATION DU BILAN ELEVAGE 2020 : Les marchés des produits laitiers, carnés et avicoles. | FranceAgriMer - établissement national des produits de l'agriculture et de la mer. https://www.franceagrimer.fr/Actualite/Filieres/Viandes-rouges/2021/PUBLICATION-DU-BILAN-ELEVAGE-2020-Les-marches-des-produits-laitiers-carnes-et-avicoles (accessed 26 May 2021).

Galusca, B., N. Germain, and B. Estour. 2016. Maigreur et hormones de régulation de l'appétit. Médecine des Maladies Métaboliques 10(1): 22–27. doi: 10.1016/S1957-2557(16)30005-0.

Gilbert, H., J. Ruesche, N. Muller, Y. Billon, V. Begos, et al. 2019. Responses to weaning in two pig lines divergently selected for residual feed intake depending on diet. J Anim Sci 97(1): 43–54. doi: 10.1093/jas/sky416.

Gondret, F., A. Vincent, M. Houée-Bigot, A. Siegel, S. Lagarrigue, et al. 2017. A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. BMC Genomics 18(1): 244. doi: 10.1186/s12864-017-3639-0.

Halford, J.C.G., and J.A. Harrold. 2012. Satiety-enhancing products for appetite control: science and regulation of functional foods for weight management†. Proceedings of the Nutrition Society 71(2): 350–362. doi: 10.1017/S0029665112000134.

Joly, P.-B. 2021. « Situé au cœur d'un système d'enseignement supérieur, de recherche et d'innovation particulièrement riche, dans une région où l'agriculture et l'agro-alimentaire sont les premiers employeurs, le centre Occitanie-Toulouse est un acteur majeur de la science et de l'innovation ouverte, au service des transitions agroécologiques des systèmes alimentaires ». : 6.

Karolchik, D., A.S. Hinrichs, T.S. Furey, K.M. Roskin, C.W. Sugnet, et al. 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32(Database issue): D493-496. doi: 10.1093/nar/gkh103.

Korthauer, K.D., S. Chakraborty, Y. Benjamini, and R.A. Irizarry. 2017. Detection and accurate False Discovery Rate control of differentially methylated regions from Whole Genome Bisulfite Sequencing. Genomics.

Le média de l'alimentaire. 2018. Le jambon cuit absorbe 73% du total de la consommation française. Les Marchés : le média des acheteurs et vendeurs de produits alimentaires. https://www.reussir.fr/lesmarches/le-jambon-cuit-absorbe-73-du-total-de-la-consommation-francaise (accessed 26 May 2021).

Magnen, J. 2012. Neurobiology of Feeding and Nutrition. Academic Press.

Mi, H., A. Muruganujan, and P.D. Thomas. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Research 41(D1): D377–D386. doi: 10.1093/nar/gks1118.

Raudvere, U., L. Kolberg, I. Kuzmin, T. Arak, P. Adler, et al. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Research 47(W1): W191–W198. doi: 10.1093/nar/gkz369.

Reverter, A., N.J. Hudson, S.H. Nagaraj, M. Pérez-Enciso, and B.P. Dalrymple. 2010. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics 26(7): 896–904. doi: 10.1093/bioinformatics/btq051.

Riebler, A., M. Menigatti, J.Z. Song, A.L. Statham, C. Stirzaker, et al. 2014. BayMeth: improved DNA methylation quantification for affinity capture sequencing data using a flexible Bayesian approach. Genome Biol 15(2): R35. doi: 10.1186/gb-2014-15-2-r35.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, et al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7): e47–e47. doi: 10.1093/nar/gkv007.

Ross-Innes, C.S., R. Stark, A.E. Teschendorff, K.A. Holmes, H.R. Ali, et al. 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481(7381): 389–393. doi: 10.1038/nature10730.

Shen, L., N.-Y. Shao, X. Liu, I. Maze, J. Feng, et al. 2013. diffReps: Detecting Differential Chromatin Modification Sites from ChIP-seq Data with Biological Replicates. PLOS ONE 8(6): e65598. doi: 10.1371/journal.pone.0065598.

Soleimani, T., S. Hermesch, and H. Gilbert. 2021. Economic and environmental assessments of combined genetics and nutrition optimization strategies to improve the efficiency of sustainable pork production. J Anim Sci 99(3). doi: 10.1093/jas/skab051.

Soneson, C., M.I. Love, and M.D. Robinson. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4: 1521. doi: 10.12688/f1000research.7563.1.

Vilain, C. 2016. Épigénétique et cancer. Planet-Vie. https://planet-vie.ens.fr/thematiques/sante/pathologies/epigenetique-et-cancer (accessed 28 April 2021).

Annexes

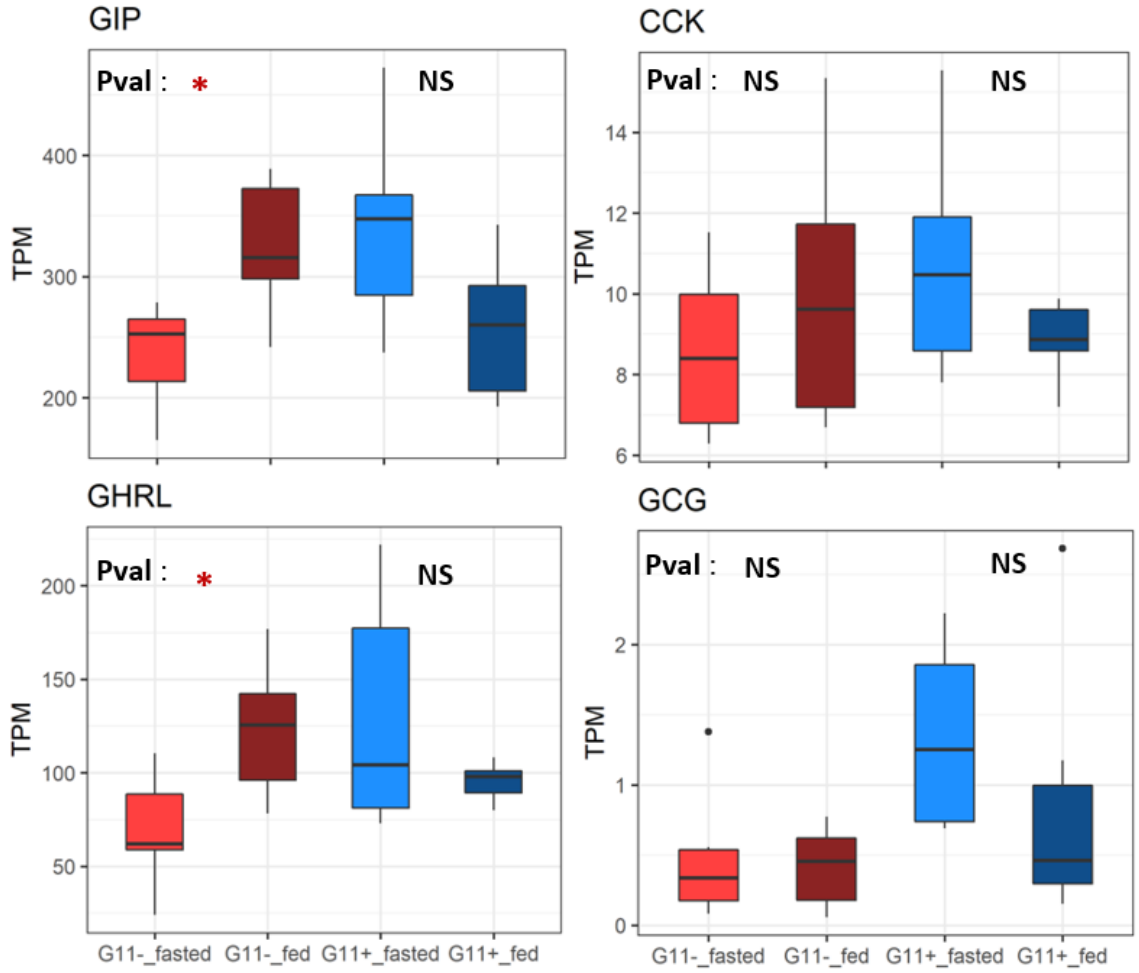

Figure 8 Représentation de l'expression différentielle chez les hormones de la régulation de la satiété. GIP augmente en expression chez les G11- de manière significative et en perd chez les G11+ pendant la prise alimentaire mais pas significativement. Un test de student sur la significativité d'expression a été réalisé sur les valeurs d'expression des hormones de régulations : les G11—fed VS les G11-_fasted et les G11+_fed VS les G11+_fasted, seules les hormones GIP et GHRL ont une expression différentielle significative (Pval = 0.014 pour GIP et pval = 0.019 pour GHRL) en prenant la valeur $\alpha = 0.05$.

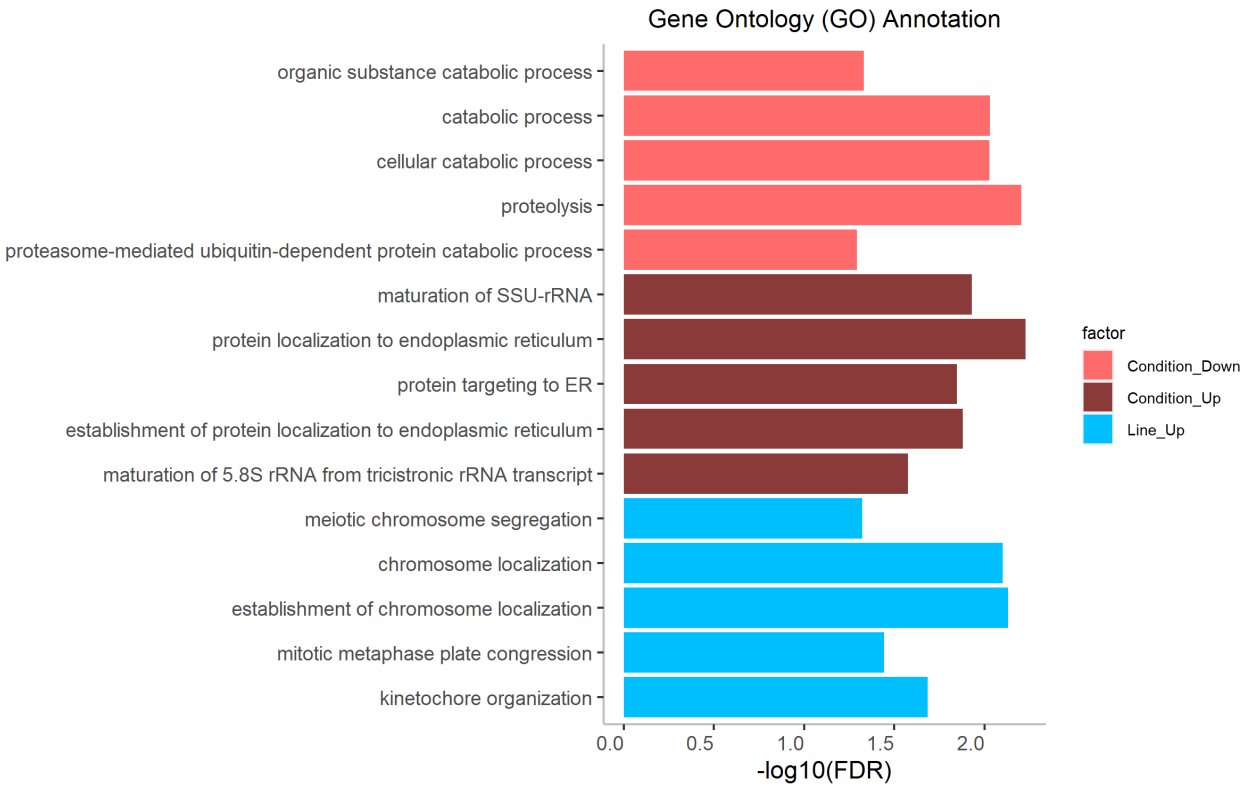

Figure 9 : Processus biologiques enrichis dans la les listes des gènes DE : L'annotation ontologiques des gènes différentiellement exprimés réalisé via PantherDB nous montre les processus biologiques les plus présents en fonction de la –log(FDR). En corail, les gènes DE avant la prise alimentaire, en rouge foncé, les gènes DE après la prise alimentaire et en bleu, les gènes DE chez les G11+. Nous ne trouvons aucun gène DE statistiquement significatif chez les G11-.

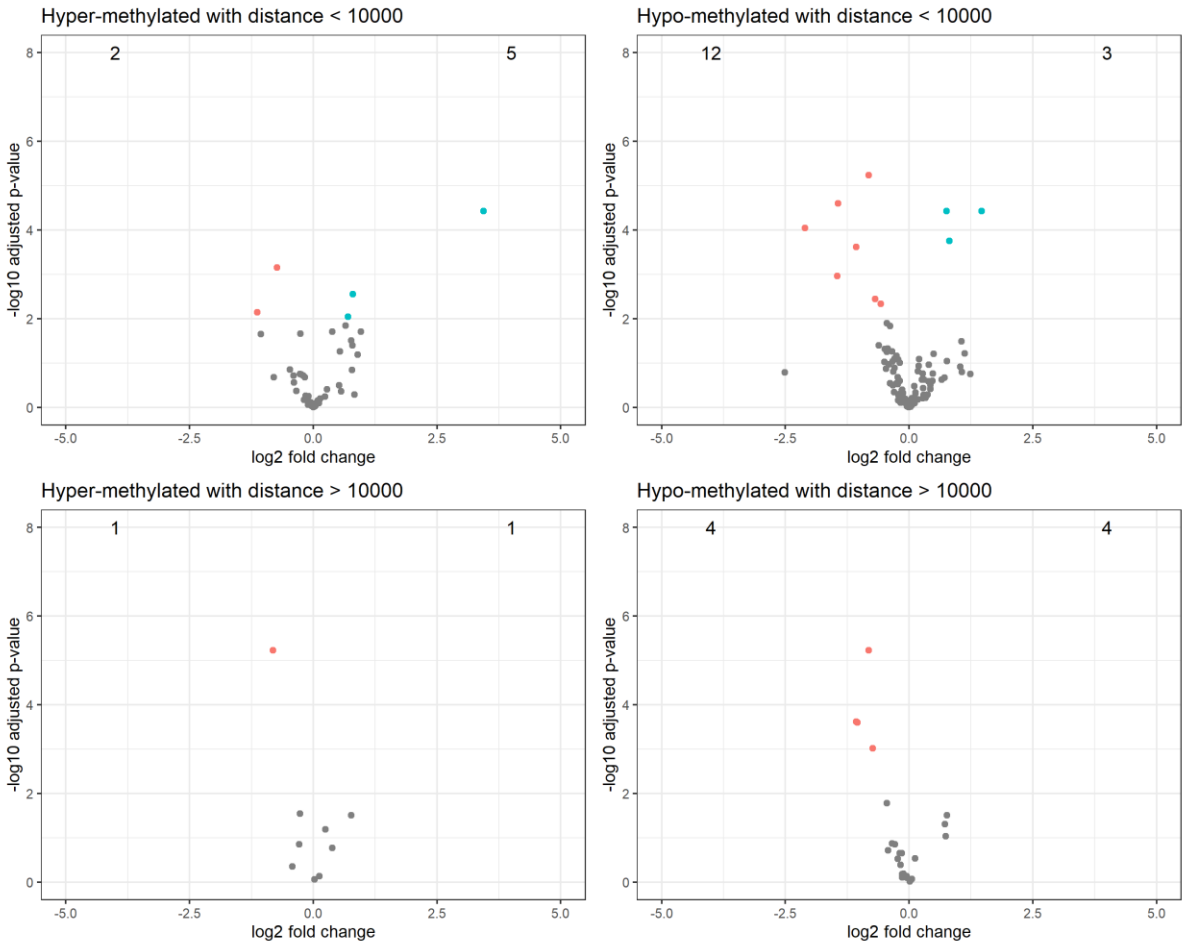

Figure 10 : Représentation des gènes up_régulé(bleu), down_régulé(rouge,) ou à expression stable(gris) les plus proches des DMRs identifiés avec DMRseq avec une distance inférieure ou supérieure à 10KB : La majorité des gènes Différentiellement méthylés ne sont pas DE (points gris), une poignée de gènes différentiellement méthylées qui sont des gènes DE sont à une distance inférieure à 10kb du DMR. Chaque point représente un gène associé à un DMR. Volcano plot des expressions différentielle selon la lignée : axe x – log2 fold change, axe y - - log10(pvalue ajusté).