



HAL
open science

Caractérisation fonctionnelle d'une région impliquée dans la croissance et l'adiposité chez le porc

Frederique Viguie, Katia Fève, Noémien Maillard, Julie Demars

► **To cite this version:**

Frederique Viguie, Katia Fève, Noémien Maillard, Julie Demars. Caractérisation fonctionnelle d'une région impliquée dans la croissance et l'adiposité chez le porc. Sciences du Vivant [q-bio]. 2025. <hal-05115132>

HAL Id: hal-05115132

<https://hal.inrae.fr/hal-05115132v1>

Submitted on 16 Jun 2025

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.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Master 2 « Biologie-Santé »

➤ **Parcours « Génomique et Biotechnologies »**

Rapport de stage présenté par :

Frédérique VIGUIE

Sujet :

Caractérisation fonctionnelle d'une région impliquée
dans la croissance et l'adiposité chez le porc

Responsables du Stage :

Katia FEVE

Noémien MAILLARD

Julie DEMARS

INRAE, CS 52627 24 Chemin de Borde Rouge
31326 CASTANET-TOLOSAN FRANCE

Juin 2025

Remerciements

Je tiens à remercier tout particulièrement mes encadrants de stage, Katia FEVE et Noémien MAILLARD, pour leur accompagnement et leur soutien constants tout au long des ces derniers mois, et notamment durant ces dernières semaines. Je les remercie également pour le temps qu'ils ont consacré à la relecture et à la correction de mon rapport.

Katia, je te remercie de m'avoir partagé ton expérience et tes conseils qui m'ont permis de progresser lors des manipulations, ainsi que pour la confiance que tu m'as accordée.

Noémien, merci pour ta patience et pour m'avoir permis de découvrir un peu plus ton monde, qui reste tout de même encore un peu obscure pour moi. Je te souhaite une bonne réussite pour la fin de ta thèse !

Ensuite, j'aimerais remercier Julie DEMARS pour son implication et ses corrections. Merci à toi !

Merci également à Anne-Sophie KIRSTETTER et Mathilde PERRET pour toutes nos discussions très agréables. Et bon courage Mathilde pour ta soutenance de thèse !

Je souhaite aussi remercier tous les membres de l'équipe GENESIS pour leur bienveillance et leur bonne humeur quotidienne.

Enfin, je tiens également à remercier l'ensemble de l'unité GenPhySE pour leur accueil.

Droits d'auteurs

Cette création est mise à disposition selon le Contrat :

« **Attribution-Pas d'Utilisation Commerciale-Pas de modification 3.0 France** »

disponible en ligne :

<http://creativecommons.org/licenses/by-nc-nd/3.0/fr/>



Résumé

Le programme PORQTL, mis en place par l'INRA en 1991, a pour objectif d'identifier des régions génomiques associées à des caractères quantitatifs (QTL) d'intérêt agronomique chez le porc. Un QTL, associé à des caractères de croissance et d'adiposité, a été localisé sur le bras long du chromosome 1 porcin, en croisant des verrats Large White et des truies Meishan. L'intervalle génomique réduit à 293 kb ne contient plus que trois gènes codants (*HMCN2*, *ASS1* et *FUBP3*). Toutefois, d'après des études bibliographiques, aucun de ces trois gènes ne semblent constituer un bon candidat fonctionnel. De plus, aucun des variants identifiés lors du séquençage de la région d'intérêt ne semblent être un variant candidat pour le QTL. L'étude s'est donc orientée vers l'hypothèse d'une mutation régulatrice, localisée dans cette région, pouvant affecter l'expression d'un gène candidat situé en dehors de l'intervalle, notamment par l'intermédiaire de l'organisation tridimensionnelle du génome. Pour explorer cette possibilité, l'expression des gènes a été mesurée par RT-qPCR et l'accessibilité de la chromatine a été étudiée par ATAC-seq.

Mots-clés : ATAC-seq, RT-qPCR, QTL, chromosome 1, porc, adiposité, croissance

Summary

The PORQTL programme, initiated by INRA in 1991, aims to determine genomic regions associated with quantitative traits (QTL) of agronomic interest in pigs. A QTL, associated with growth and adiposity traits, was located on the long arm of the chromosome 1 of the pig genome, by crossing between Large White boars and Meishan sows. The genomic interval was reduced to 293 kb and now contains just three coding genes (*HMCN2*, *ASS1* and *FUBP3*). However, according to the literature, none of these three genes appear to be a good functional candidate. Furthermore, none of the variants identified during sequencing of the region of interest appear to be candidate variants for the QTL. The study therefore focused on the hypothesis that a regulatory mutation located in this region could affect the expression of a candidate gene located outside the interval, notably via the three-dimensional organisation of the genome. To explore this possibility, gene expression was measured by RT-qPCR and chromatin accessibility was studied by ATAC-seq.

Keywords : ATAC-seq, RT-qPCR, QTL, chromosome 1, pig, adiposity, growth

Table des abréviations

- A -

ADNc : ADN complémentaire

ASS1 : Argininosuccinate synthase

ATAC-seq : Assay for Transposase Accessible Chromatin followed by high throughput sequencing

- B -

BMP15 : Bone Morphogenetic Protein 157

- C -

ChIP-seq : Chromatin ImmunoPrecipitation followed by high throughput sequencing

cM : centimorgan

Ct : cycle seuil

CTCF : CCCTC-binding factor

- D -

DNase-seq : DNase I hypersensitive sites sequencing

- F -

FAANG : Functional Annotation of Animal Genomes

FNBPI : Formin Binding protein 1

FT : facteur de transcription

FTO : Fat mass and obesity associated protein

FUBP3 : Far Upstream Element Binding Protein 3

- G -

GDF9 : Growth Differentiation Factor-9

GPR107 : G protein-coupled receptor 107

- H -

HMCN2 : Hemicentin 2

HPRT : Hypoxanthine guanine Phosphoribosyl Transferase

- I -

IA : Intelligence Artificielle

IGV : Integrative Genomics Viewer

INRA : Institut National de la Recherche Agronomique

INRAE : Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

IRX3 : Iroquois Homeobox 3

- L -

LW : Large White

- M -

MgCl₂ : Chlorure de Magnésium

MLPH : Melanophilin

MS : Meishan

Mtec : millions de tonnes équivalents carcasse

- N -

NGS : Next-Generation Sequencing

NCSI : Neuronal Calcium Sensor 1

- P -

PBS : Phosphate-buffered saline

- Q -

QTL : Quantitative Trait Loci

- R -

Rec : recombinant

RNA-seq : RNA sequencing

RPL32 : Ribosomal Protein L32

RT-qPCR : Reverse Transcription quantitative real-time PCR

- S -

SPRI : Solide Phase Reversible Immobilization

SSC1 : *Sus scrofa* chromosome 1

- T -

TAE : Tris Acetate EDTA

TBP1 : TATA-box binding protein

TES : Transcription end sites

TOP2B : DNA topoisomerase II beta

TSS : Transcription start sites

- Autre -

3D : 3 dimensions

Table des matières

Introduction	1
Cadre socio-économique.....	1
Détection de régions génomiques d'intérêt – le programme PORQTL.....	2
Analyses génétiques d'un QTL localisé sur le chromosome 1 porcin.....	3
Gènes candidats fonctionnels localisés dans l'intervalle du QTL d'intérêt.....	3
Régulation de l'expression des gènes.....	4
Objectif.....	6
Travail expérimental	8
Matériels et Méthodes.....	8
Animaux.....	8
Échantillons biologiques.....	8
Synthèse des ADNc.....	8
PCR en point final.....	9
PCR quantitative en temps réel.....	9
Statistiques.....	10
Préparation et isolement des noyaux.....	11
Préparation des bibliothèques ATAC-seq.....	11
Séquençage.....	13
Traitements des données ATAC-seq.....	13
Résultats.....	14
Partie 1 : Etude de l'expression des 6 gènes d'intérêt.....	14
Partie 2 : Etude de l'accessibilité de la chromatine.....	16
Discussion/conclusion	22
Discussion.....	22
Conclusion – Perspectives.....	25
Bibliographie	26

Table des illustrations

Figure 1 : Production et consommation mondiale de viande porcine

Figure 2 : Schéma du chromosome 1 d'un individu recombinant/Meishan

Figure 3 : Schéma de la région étudiée

Figure 4 : Schéma de l'organisation et de la régulation fonctionnelle du génome

Figure 5 : Programme d'amplification du QuantStudio 6

Figure 6 : Protocole Active Motif d'extraction des noyaux

Figure 7 : Préparation des librairies

Figure 8 : Courbe de dissociation pour le gène HPRT

Figure 9 : Mesure l'efficacité des couples d'amorces

Figure 10 : Analyse de la stabilité des gènes de référence réalisée avec geNorm

Figure 11 : Courbe de dissociation pour le gène FUBP3

Figure 12 : Expression relative des 6 gènes d'intérêt dans les 3 tissus

Figure 13 : Contrôle des données pour le foie

Figure 14 : Observation au microscope des noyaux avant et après optimisation

Figure 15 : Observation du culot nucléaire après gradient d'iodixanol

Figure 16 : Profils des librairies obtenues au Fragment Analyzer

Figure 17 : Qualité des lectures de séquençage selon les échantillons et les tissus

Figure 18 : Analyse des lectures de séquençage pour chaque échantillon et chaque tissu

Figure 19 : Représentation IGV focalisée sur le gène ASS1 pour chaque échantillon

Figure 20 : Contrôles qualité attendu de l'analyse bio-informatique d'ATAC-seq

Figure 21 : Représentation IGV d'une région du chromosome 2

Table des tableaux

Tableau 1 : Séquences des couples amorces pour les gènes cibles et de références

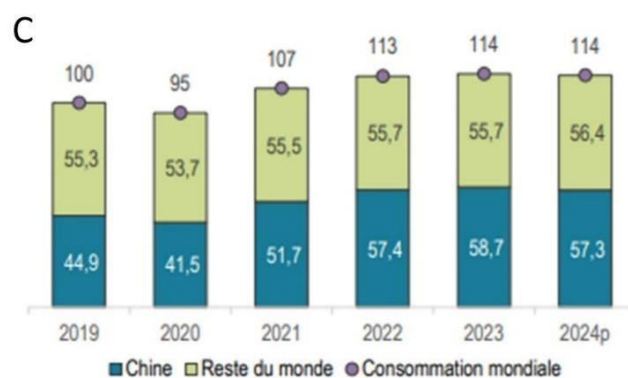
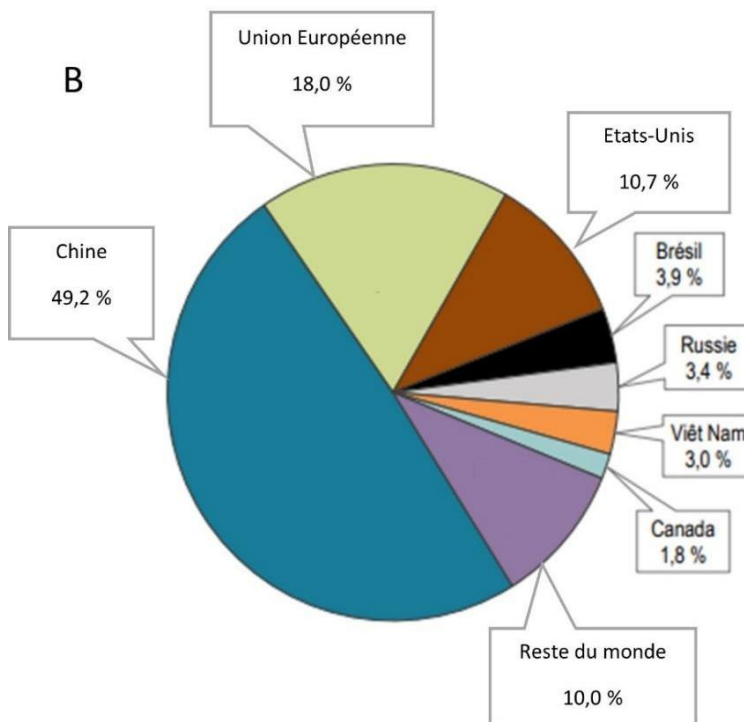
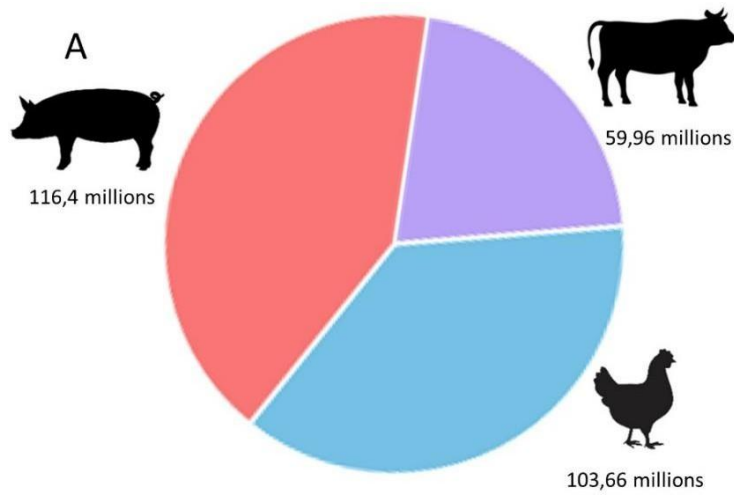


Figure 1 : Production et consommation mondiale de viande porcine. A. Production mondiale de viande (2023-2024) en millions de tonnes (d'après <https://www.fas.usda.gov/data/production/commodity-group/meat>). B. Les principaux pays producteurs de viande porcine en 2023. C. Variation de la consommation de viande porcine en millions de tonnes équivalent carcasse (Mtec) dans le monde, entre 2019 et 2023. 2024p correspond à la prévision de l'USDA pour 2024 (d'après Zarnovican, 2024).

Introduction

Cadre socio-économique

La viande porcine domine la consommation mondiale de viande (**Figure 1A**). En 2023, les trois principaux producteurs de porc étaient la Chine avec une production de 49,2%, suivie par l'Union européenne (18,0%) et les Etats-Unis (10,7%) (**Figure 1B**). La Chine est également le premier consommateur de viande porcine avec une consommation de 58,7 millions de tonnes équivalents carcasse (Mtec) en 2023, soit plus de la moitié de la consommation mondiale (114 Mtec) (**Figure 1C**) (FranceArgiMer. 2025). En Europe, la France s'est classée comme étant le troisième pays producteur de porc avec 36% de sa production exportée dans le monde en 2023. De plus, la consommation française globale s'élevait à 2,1 Mtec, soit environ 30,5 kg équivalent carcasse par habitant (Zarnovican. 2024). Afin de pouvoir satisfaire durablement les besoins, la productivité doit être optimisée en passant par une sélection des lignées porcines, en choisissant les meilleurs reproducteurs selon des critères génomiques et phénotypiques (Robert-Granié et al. 2011). Au cours des dernières décennies (1990-2010), les critères privilégiés par les sélectionneurs étaient principalement des caractères de production et de reproduction, notamment le taux de muscle. Tout en favorisant le développement musculaire, l'objectif était d'obtenir une viande plus maigre, permettant ainsi de répondre à la demande des consommateurs pour des produits plus sains. Cependant, le tissu adipeux joue un rôle important dans la qualité gustative et les caractéristiques nutritionnelles de la viande (Schwob et al. 2020). Il joue également un rôle fondamental dans la robustesse des animaux, en particulier chez les truies lors de la reproduction. En effet, une certaine quantité de graisse corporelle améliore la résilience pendant la reproduction car elle fournit des réserves d'énergie essentielles au cours de la gestation et l'allaitement (Quéméneur et al. 2023). L'introgession de gènes codant pour des caractères de gras permettrait donc d'assurer une carrière plus longue des truies, répondant

aussi au développement durable qui est un nouveau critère demandé par les consommateurs.

Détection de régions génomiques d'intérêt – le programme PORQTL

Les caractères de croissance et de gras sont des caractères complexes. Leur expression est régulée par plusieurs gènes à effet faible sur le phénotype (masse musculaire, taille, etc), ainsi que par des facteurs environnementaux. Chaque région génomique impliquée dans la variabilité de ces caractères est appelé QTL (Quantitative Trait Loci ou Locus à effet quantitatif) (Abiola et al. 2003). La somme des effets de toutes les mutations causales de tous les QTL pour un caractère donné, représente alors la composante génétique responsable d'un phénotype.

Dans les années 1990, l'Institut National de la Recherche Agronomique (INRA, actuellement INRAE depuis le 1^{er} janvier 2020) a lancé le programme PORQTL dans le but d'identifier des QTL pour un grand nombre de caractères d'intérêt pour la filière porcine, tels que la composition de la carcasse, le taux de viande dite « maigre », l'épaisseur du lard dorsal... (Bidanel and Milan 2000). Un dispositif a été mis en place, dans lequel des croisements ont été réalisés entre des verrats Large White (LW) et des truies Meishan (MS), qui sont deux races porcines présentant des phénotypes très différents. La race Large White est une race européenne. Elle est très utilisée dans la production de viande porcine pour ses caractères de croissance musculaire. Elle est également connue pour sa capacité à produire un rendement élevé en viande « maigre » (Taylor et al.). La race Meishan, originaire de Chine, est quant à elle reconnue pour ses caractéristiques de maturité précoce et sa forte prolificité, mais aussi pour sa qualité de viande « grasse » (Legault et al. 1983). Les premières analyses génétiques de détection de QTL dans ce protocole expérimental de grande ampleur ont permis d'identifier, entre autres, une région associée à des caractères de croissance et d'adiposité sur le bras long du chromosome 1 (SSC1) (Bidanel et al. 2001).

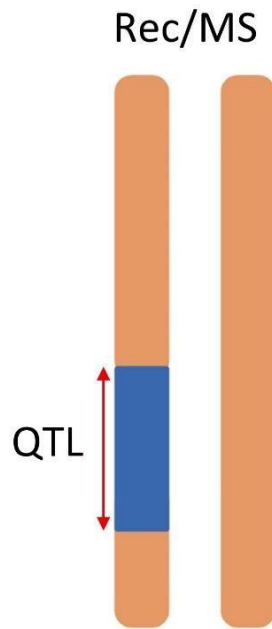


Figure 2 : Schéma du chromosome 1 d'un individu recombinant/Meishan. Bleu : allèle Large White, orange : allèle Meishan. La localisation du QTL est indiquée par une double flèche rouge.

Sscrofa 11.1

1:270,004,000-270,700,000

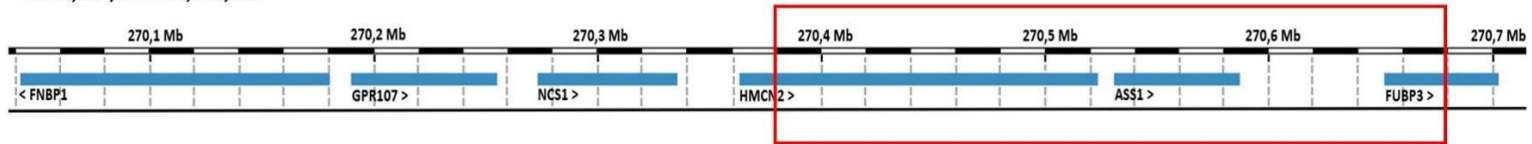


Figure 3 : Schéma de la région étudiée. En rouge est encadré la région du QTL du chromosome 1 (270,380,000-270,680,000).

Analyses génétiques d'un QTL localisé sur le chromosome 1 porcin

Les travaux de Milan *et al* ont permis de mettre en évidence un effet favorable et récessif de l'allèle Meishan pour le gras (Milan et al. 2002). En effet, la présence de cet allèle est liée à une plus grande adiposité, se traduisant par une épaisseur de lard dorsal plus importante.

Cette première approche a ainsi mis en évidence un QTL sur le bras long du chromosome 1 dans un intervalle de 40cM, affectant les caractères de croissance et d'adiposité.

Des approches génétiques complémentaires ont permis de réduire l'intervalle du QTL, pour limiter le nombre de gènes candidats à étudier et potentiellement faciliter l'identification de la mutation causale, responsable de l'effet du QTL. Pour ce faire, des rétro-croisements ont été réalisés, conduisant à l'obtention d'individus recombinants/Meishan (Rec/MS). Ces individus sont hétérozygotes (LW/MS) pour le QTL et homozygotes MS autour du QTL (**Figure 2**). Ainsi, la région génomique du QTL a progressivement été significativement réduite de plusieurs mégabases à quelques centaines de kilobases (communication personnelle, Fève). Les analyses de séquençage nouvelle générations (NGS) d'individus d'intérêt, ont permis d'identifier plusieurs centaines de variants dans l'intervalle du QTL. Cependant, la majorité de ces variants est localisée dans des séquences non codantes (Demars et Mourad. 2023).

Gènes candidats fonctionnels localisés dans l'intervalle du QTL d'intérêt

Aujourd'hui, l'intervalle génomique fait 293 kb et comporte 3 gènes codants (*HMCN2*, *ASS1* et *FUBP3*) (Fève) (**Figure 3, encadré rouge**). Cependant, l'étude bibliographique n'a pas permis de mettre en évidence un gène candidat fonctionnel. En effet, l'Hemicentin 2 (*HMCN2*) appartient à la famille des fibulines, qui sont des protéines de la matrice extracellulaire, et joue un rôle dans l'adhésion tissulaire ainsi que la migration cellulaire (Welcker et al. 2021). Le gène *ASS1* (Argininosuccinate synthase) quant à lui, code une enzyme essentielle dans le cycle de l'urée. Chez l'Homme, une mutation du gène provoque

le trouble du cycle de l'urée aussi appelé citrullinémie (Rabinovich et al. 2015). Enfin, le gène *FUBP3* (Far Upstream Element Binding Protein 3) code une protéine de liaison à l'ADN simple brin impliquée dans la régulation transcriptionnelle (Davis-Smyth et al. 1996).

En revanche, les gènes *GPR107* et *FNBPI*, situés en amont de l'intervalle, pourraient constituer de bons candidats fonctionnels (**Figure 3**), à l'inverse du gène *NCSI* (Neuronal Calcium Sensor 1), localisé entre *GPR107* et *HMCN2* (**Figure 3**), code une protéine de liaison au calcium neuronal (Fischer et al. 2021), ce qui rend moins probable son implication directe dans le mécanisme ciblé. Le gène *GPR107* (G protein-coupled receptor 107) joue un rôle dans l'adipogenèse en codant un récepteur de la neuronostatine, qui se trouve impliquée dans la formation de tissu adipeux blanc où l'énergie est stockée sous forme de lipides (Krażek et al. 2024). Le gène *FNBPI* (Formin Binding protein 1) code une protéine de liaison à la formine qui participe à la mise en place du réseau d'actine (Aspenström 2010). Le gène *GPR107* pourrait donc avoir un lien fonctionnel pour le caractère d'adiposité et *FNBPI* pour le caractère de croissance. Même si ces gènes ne sont pas dans l'intervalle, il est connu que des mécanismes de régulation des gènes peuvent se faire à longue distance. Ces résultats orientent donc l'étude vers une potentielle mutation régulatrice, susceptible d'affecter l'expression d'un gène candidat *via* notamment l'organisation en 3 dimensions (3D) du génome. Une étude chez la souris a montré que des variants non codants associés à l'obésité au sein du gène *FTO*, modulent l'expression du gène *IRX3* en agissant à distance par des boucles chromatiniennes (Smemo et al. 2014). Il se peut donc que des variants présents dans le QTL régulent ces gènes, bien qu'ils soient en dehors de l'intervalle (Umlauf and Mourad 2019).

Régulation de l'expression des gènes

Les caractères complexes sont souvent contrôlés par des variants non-codants (plus de 90%) qui ont un impact sur la régulation de l'expression génique, rendant difficile

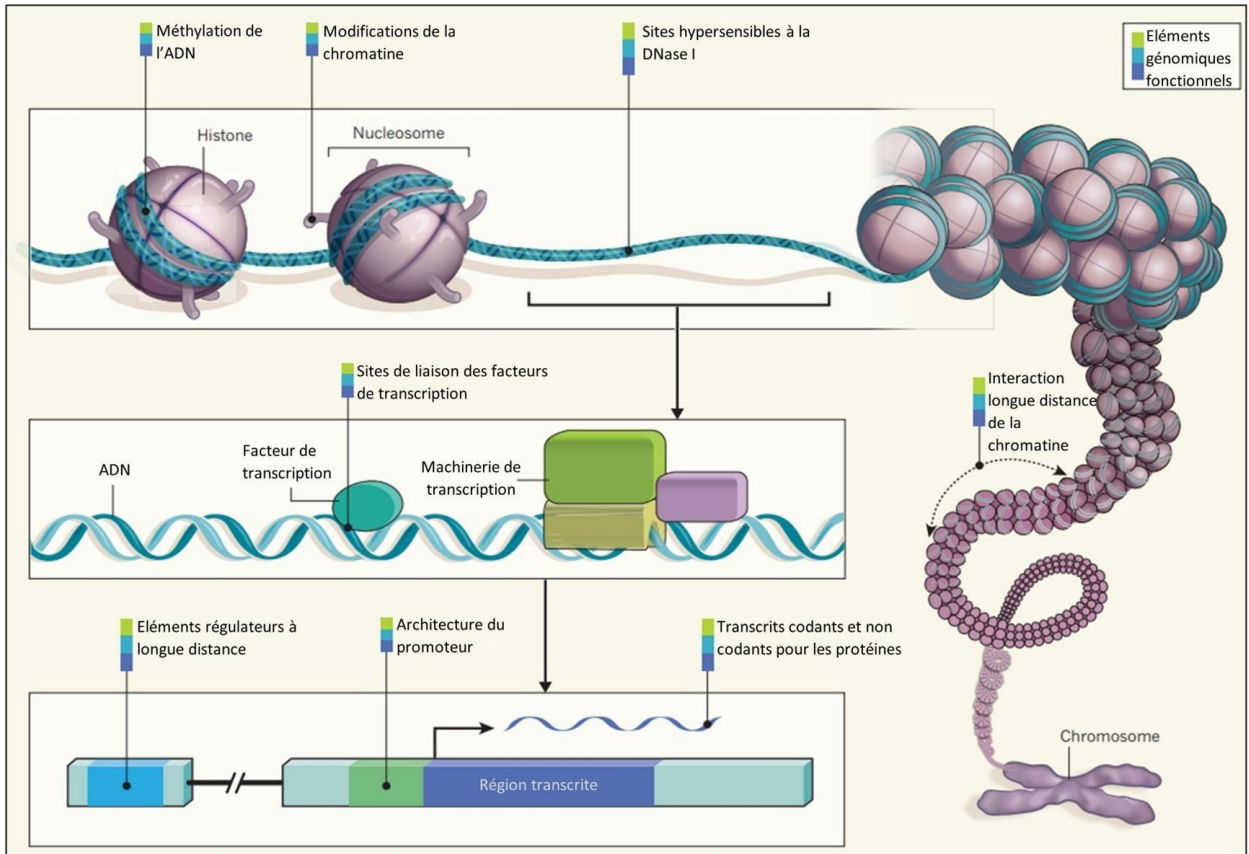


Figure 4 : Schéma de l'organisation et de la régulation fonctionnelle du génome. Modifié d'après Ecker et al. 2012.

l'identification des mutations causales (Umlauf and Mourad 2019). Le consortium FAANG (Functional Annotation of Animal Genomes) a été créé pour annoter fonctionnellement les génomes d'animaux d'élevage et identifier les éléments régulateurs impliqués dans la régulation de l'expression des gènes (Giuffra et al. 2019). La régulation de l'expression des gènes est affectée notamment par le niveau de compaction de la chromatine : on parlera de « régulation de la chromatine ». La chromatine est une structure composée d'ADN et de nucléosomes. Ces derniers sont constitués d'un cœur d'histones autour desquelles s'enroule 146 pb d'ADN (Lee and Mahadevan 2009). La chromatine « ouverte », permettant aux facteurs de transcription (FT) et à l'ARN polymérase de se lier aux promoteurs des gènes, est appelée euchromatine tandis que la chromatine « fermée » est nommée hétérochromatine (**Figure 4**). Les régions régulatrices de la chromatine incluent les éléments *cis*-régulateurs. Ces éléments sont des séquences ayant une action sur un gène à proximité. Elles permettent de réguler l'expression des gènes en recrutant des protéines, modulant l'accessibilité de la chromatine (Wang et al. 2007; Lu et al. 2018).

Afin d'étudier la structure et la dynamique de la chromatine, différentes méthodes d'analyse de son accessibilité ont été développées, ces dernières années. Parmi celles-ci, la DNase-seq (hypersensibilité à la DNase I suivi de séquençage haut débit) a été l'une des premières à être mise en place en 2006. Cependant, cette technique est longue et nécessite une grande quantité de matériel, c'est pourquoi elle a été remplacée par l'ATAC-seq (Assay for Transposase Accessible Chromatin sequencing). Cette dernière est de plus en plus utilisée compte tenu de sa rapidité d'exécution et de son faible besoin en nombre de cellules. L'ATAC-seq est une technique qui permet de déterminer l'état de compaction de la chromatine. Cette méthode repose sur l'utilisation d'une enzyme, la transposase Tn5 hyperactive, qui fragmente l'ADN au niveau des régions ouvertes de la chromatine et qui insère des adaptateurs essentiels

pour la préparation des bibliothèques, pour le séquençage haut débit (Buenrostro et al. 2013). Cette méthode permet donc d'identifier les régions de chromatine accessibles à l'échelle du génome entier

La méthode de ChIP-seq (Chromatin ImmunoPrecipitation sequencing) est également fréquemment utilisée. Elle permet une analyse plus fine de l'étude de la régulation de la chromatine et de l'expression génique. Cette technique permet d'identifier spécifiquement, à l'échelle du génome entier, les sites de liaison de protéines à l'ADN, tels que les FT ou les modifications post-traductionnelles sur les histones. Cette approche repose sur l'utilisation d'anticorps ciblant spécifiquement les protéines d'intérêts (Lee and Mahadevan 2009). A titre d'exemple, la ChIP-seq peut aussi permettre d'identifier les sites d'ancrage des boucles chromatiniques en utilisant un anticorps ciblant la protéine CTCF. Cette protéine rapproche deux régions d'ADN, permettant ainsi l'interaction entre des éléments régulateurs éloignés qui peuvent influencer la régulation de l'expression des gènes (Umlauf and Mourad 2019).

Ces deux techniques sont complémentaires et permettent d'analyser la régulation de l'expression des gènes à des niveaux différents. L'ATAC-seq identifie les régions accessibles de la chromatine, celles ouvertes et donc potentiellement actives. De son côté, la ChIP-seq met en évidence les interactions entre l'ADN et certaines protéines (facteurs de transcription ou modifications d'histones) grâce à l'utilisation d'anticorps spécifiques. Ainsi, l'ATAC-seq est une approche sans a priori qui fournit une vision globale des régions ouvertes de la chromatine afin d'identifier un gène effecteur par exemple, alors que la ChIP-seq apporte une information plus précise sur l'implication de facteurs ou marques d'histones spécifiques.

Objectifs

Chez le porc, un QTL associé à la croissance et à l'adiposité a été identifié sur le chromosome 1. Comme suggéré précédemment, (*i*) aucun variant génétique potentiellement.

causal n'est localisé dans les séquences codantes des gènes candidats positionnels et (ii) des gènes à proximité mais en amont du QTL peuvent être de bons candidats fonctionnels. L'hypothèse actuelle est qu'un variant régulateur localisé dans la région du QTL influencerait l'expression de l'un de ces gènes *via* des mécanismes de régulation à longue distance.

L'objectif de ce stage a été d'approfondir la compréhension des mécanismes de régulation dans la région d'intérêt, chez des individus génétiquement opposés au niveau du QTL (Rec/MS et MS/MS). Dans un premier temps, l'expression des 6 gènes situés dans et en amont de l'intervalle a été quantifiée par RT-qPCR, afin d'identifier d'éventuelles différences d'expression entre les deux génotypes. Par la suite, l'accessibilité de la chromatine a été évaluée par ATAC-seq pour identifier les régions actives ou non de la chromatine, et ainsi confirmer les résultats obtenus en qPCR. Sur la partie ATAC-seq, mon travail a consisté à optimiser le protocole de préparation des échantillons avant le séquençage et à l'analyse bio-informatique.

Travail expérimental

Matériels et Méthodes

Animaux

20 animaux : 10 animaux génétiquement Rec/MS pour le QTL et 10 individus MS/MS, ont été abattus à 30 kg à l'abattoir de l'unité expérimentale du Magneraud (UE GenESI, Elevages Porcins Innovants, (<https://doi.org/10.15454/1.5572415481185847E12>)).

Échantillons biologiques

Pour chacun des 20 individus, 3 tissus ont été prélevés : la longe (muscle), le tissu adipeux et le foie. Chez le porc, la longe correspond au muscle dorsal ce qui en fait un tissu d'intérêt pour l'étude des caractères liés à la croissance. Le tissu adipeux et le foie, quant à eux, constituent des tissus cibles pour le caractère d'engraissement. En effet, chez le porc, la lipogenèse est principalement localisée dans le tissu adipeux mais cette dernière peut être également localisée dans le foie suivant le stade de développement (Mourot et al. 1999). Les tissus ont été prélevés directement à l'abattoir de l'unité expérimentale du Magneraud (UE GenESI, Elevages Porcins Innovants). Au moment du prélèvement, les tissus ont été découpés en morceaux d'environ 1 cm³, puis rapidement congelés dans de l'azote liquide et conservés à -80°C. Des morceaux de tissus congelés ont ensuite réduits en poudre, puis stockés à -80°C.

Synthèse des ADNc

Les ARN totaux ont été extraits à partir de tissus mis en poudre puis transcrits en ADN complémentaires (ADNc), à l'aide du kit SuperScript™ First-Strand Synthesis System for RT-PCR (Invitrogen®), suivant les recommandations du fournisseur, décrites ci-dessous. 1 µg d'ARN total a été ajouté dans un volume de 10 µL d'eau, afin que tous les échantillons soient ajusté au même volume final. 1 µL d'oligonucléotides « random » à 100 µM (Sigma oligos) ainsi que 1 µL de dNTPs (10 mM) (Eurobio) ont été ajoutés. Pour dénaturer les ARN totaux,

ces derniers sont incubés à 65°C pendant 5 min. Ils sont ensuite immédiatement refroidis sur glace. Ensuite, 4 µL de tampon First Strand Buffer 5X, 2 µL de DTT (0,1 M), 1 µL de RNAsine et 1 µL de reverse transcriptase Superscript II sont ajoutés au mélange contenant les ARN totaux. Pour l'étape de synthèse des ADNc, le mélange réactionnel est ensuite incubé 10 min à 25°C, puis 50 min à 42°C et 15 min à 72°C. La synthèse d'ADNc est ensuite vérifiée par la réalisation d'une PCR en point final avec un couple d'amorce d'un gène de référence et un dépôt sur agarose. Les échantillons d'ADNc obtenus sont dilués au 1/10^{ème} et conservés à -20°C jusqu'à utilisation.

PCR en point final

Afin de vérifier la synthèse d'ADNc effectuée précédemment, une PCR des ADNc est réalisée avec le couple d'amorces *HPRT*, dans un volume final de 15 µL contenant du tampon PCR 1X avec 3 mM de MgCl₂ inclus (PCRBIO Classic Taq, Eurobi), 200 µM de dNTP (Eurobio), 0,25 µM d'amorces, 0,25 U d'ADN polymérase (PCRBIO Classic Taq, Eurobi) et 2 µL d'ADNc dilué au dixième. La PCR est réalisée dans un thermocycleur 2720 Thermal Cycler (Applied Biosystems) avec les conditions d'amplifications suivantes : une dénaturation initiale à 95°C pendant 5 min, 35 cycles comprenant une dénaturation de 35 sec à 94°C, une hybridation de 30 sec à une température correspondant au T_m de chaque couple d'amorces et une élongation de 20 sec à 72°C, pour finir une élongation finale à 72°C pendant 20 min. Les produits d'amplification sont analysés par électrophorèse sur gel d'agarose à 150 V. Pour cela, un gel d'agarose à 2% est préparé avec du tampon TAE 1X (Tris Acetate EDTA). Les produits PCR sont visualisés grâce au colorant fluorescent GelRed (Interchim). Pour vérifier la taille des fragments, un marqueur de taille (PCRBIO Ladder IV, Eurobio) est utilisé.

PCR quantitative en temps réel

Tableau 1 : Séquences des couples d'amorces pour les gènes cibles et de références.

Gène		Amorce foward	Amorce reverse
FNBP1	Cible	CCTGCCATAGGGACCTGTAA	GTATGACGTGGGGACGTAGC
GPR107	Cible	CACGCTGGTGTCTTTGTTT	CCGTGGTCACCTTCTTGAC
NCS1	Cible	GGAAGGCTCCAAGGCAGA	GTGTGGATGCGGAGAAAGG
HMCN2	Cible	CCTGGACGAGTGTCAAGTG	GTAGCTGCCACGGGTGTT
ASS1	Cible	CCCTCTACAACGAGGAGCTG	GAGAGATGGCAAAGTGAGG
FUBP3	Cible	ACAGCAGGTCGCTTTCTACG	TCACAGAAAACAAAGGTTCTTCA
HPRT	Référence	TACCTAATCATTATGCCGAGGATT	AGCCGTTCACTCCTGTCCAT
TBP1	Référence	AACAGTTCAGTAGTTATGAGCCAGA	AGATGTTCTCAAACGCTTCG
RPL32	Référence	TGGTTATGGGAGCAACAAGA	GGAAGAGACGTTGTGAGCA
ActineB	Référence	CACGCCATCCTGCGTCTGGA	TGGAAAACTCCGTATCTGTCTC
TOP2B	Référence	AACTGGATGATGCTAATGATGCT	TGGAAAACTCCGTATCTGTCTC

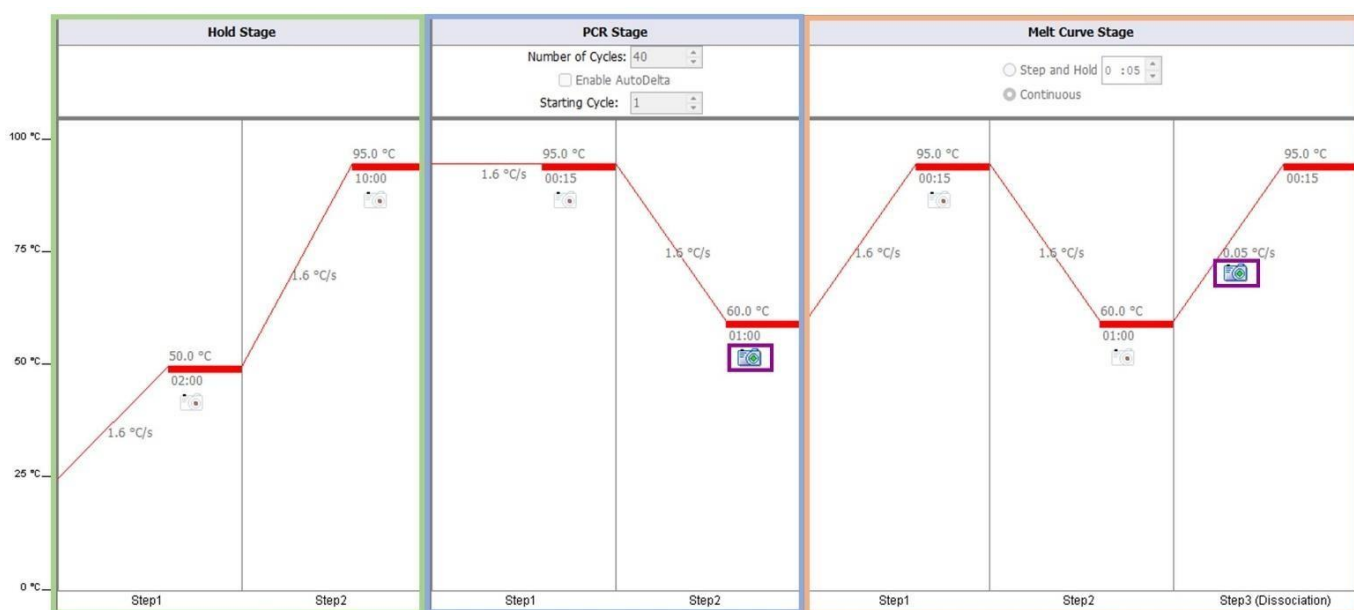


Figure 5 : Programme d'amplification du QuantStudio 6. En vert est encadré la phase d'activation et de dénaturation initiale, en bleu la réaction d'amplification et en orange l'étape de fusion. Les mesures d'intensité de fluorescence sont indiquées par un appareil photo encadré violet.

La qPCR mesure l'amplification de l'ADN à chaque cycle en détectant un signal fluorescent proportionnel à la quantité d'ADN produite, grâce à un intercalant fluorescent comme le SYBR Green. Les couples d'amorces utilisés avaient déjà été testés au sein du laboratoire. Les séquences des couples pour les gènes ciblés et les gènes de références sont indiqués dans le **Tableau 1**. Les réactions de qPCR ont été réalisées dans un volume final de 5 µL composé de 1,5 µL d'ADNc dilué, de 2,5 µL de SYBR green PCR MasterMix 2X (FischerScientific) et 150 nM de chaque amorce. La réaction d'amplification comprend 40 cycles. Chaque cycle est composé d'une étape de dénaturation à 95°C pendant 15 sec, suivie d'une étape d'hybridation/élongation pendant 1 minute à 60°C. L'intensité de fluorescence est détectée à la fin de chaque cycle d'amplification par l'appareil QuantStudio 6 (Thermo Fisher Scientific), afin de quantifier en temps réel l'ADN amplifié. A la fin de la réaction d'amplification, une étape de fusion est réalisée, durant laquelle la température monte de 60°C à 95°C par palier de 0,05°C/s (**Figure 5**). L'ADN double brin se dénature avec l'augmentation progressive de la température, faisant chuter la fluorescence, ce qui permet de tracer une courbe de dissociation. La présence d'un seul pic indique que seul le fragment d'intérêt a été produit lors de l'amplification. Cette courbe permet donc de contrôler la spécificité des produits amplifiés.

Statistiques

Par la suite, la méthode du double delta CT est utilisée pour déterminer l'expression relative des gènes. La formule est la suivante :

$$\Delta Ct = Ct_{gène} - Ct_{calibrateur}$$

$$\Delta\Delta Ct = \Delta Ct_{gène} - \Delta Ct_{gène\ référence}$$

$$QR = 2^{-\Delta\Delta Ct}$$

Une moyenne géométrique des 2 gènes de références est réalisée pour la normalisation des données. Les gènes ont ensuite été analysés à l'aide d'un test de Student bilatéral (excel, R).

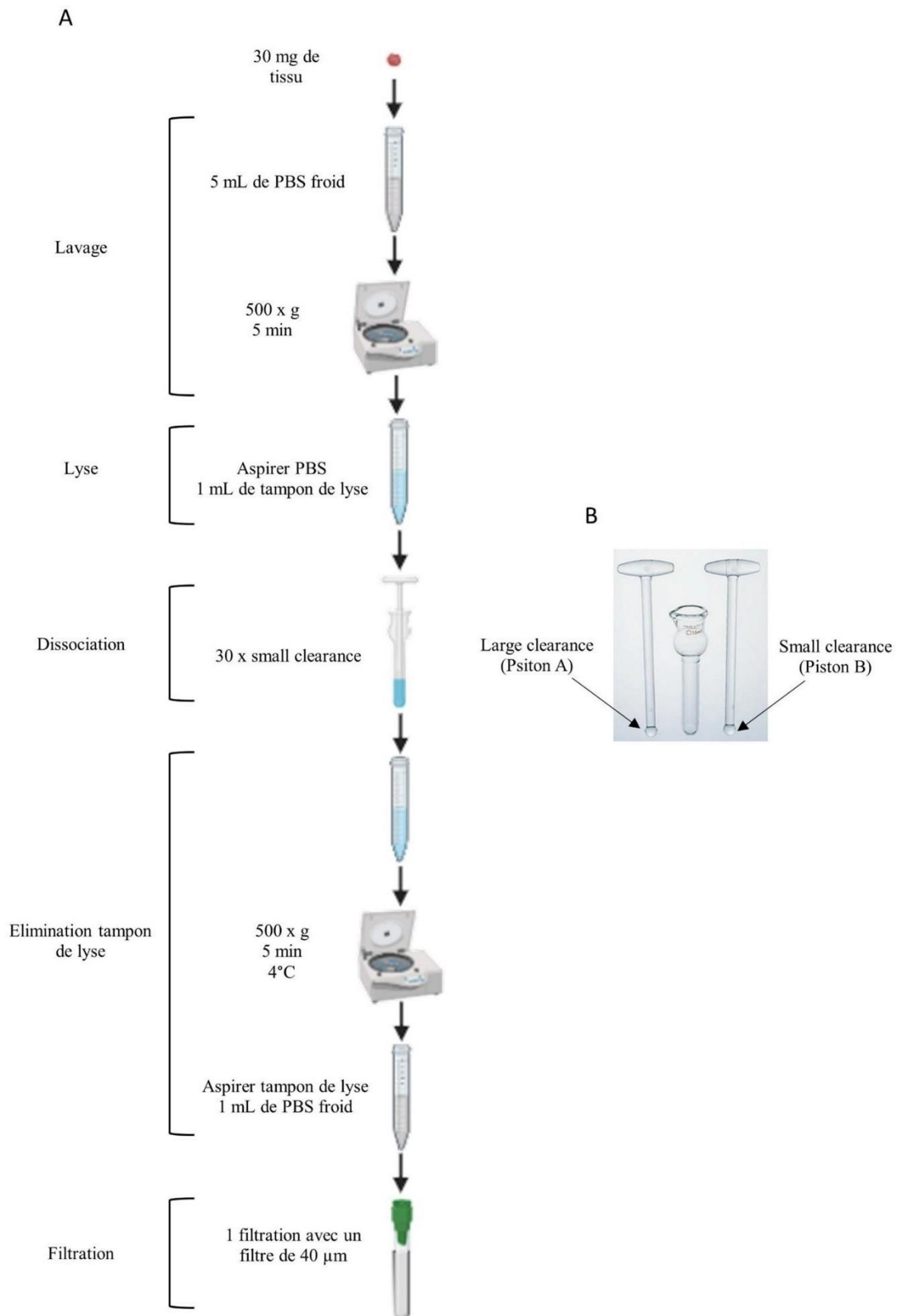


Figure 6 : Protocole Active Motif d'extraction des noyaux. **A.** Etapes du protocole Active Motif (ATAC-seq Kit #53150) sans modifications pour la préparation et l'isolement des noyaux. **B.** Représentation d'un homogénéiseur douce composé d'un cylindre en verre et de 2 pistons : large clearance (piston A) et small clearance (piston B).

Préparation et isolement des noyaux

Parmi les 20 échantillons, 2 individus Rec/MS et 2 individus MS/MS ont été sélectionnés pour les 3 tissus. Les étapes de l'isolement des noyaux décrites ci-dessous correspondent au protocole du kit Active Motif (ATAC-seq Kit #53150) sans modifications. Le protocole est résumé **Figure 6A**. Les noyaux ont été isolés à partir de morceaux congelés de foie, de tissu adipeux et de longe. Dans un premier temps, les tissus ont été hachés à l'aide d'un scalpel sur une boîte de pétri placée sur glace. Par la suite, chaque échantillon a été transféré dans 1 ml de PBS glacé, puis centrifugé à 500 g pendant 5 min à 4°C. Une fois le surnageant retiré, 1 mL de tampon de lyse a été ajouté. Les échantillons ont ensuite été transvasés dans un dounce de 1 mL, puis homogénéisés à l'aide du piston « small clearance » en effectuant 30 allers-retours. Un dounce est un homogénéiseur manuel en verre, utilisé pour dissocier délicatement des tissus ou cellules, tout en préservant l'intégrité des noyaux. Il est composé de 2 pistons de diamètres différents : un « large clearance » (piston A) pour la réduction initiale des échantillons et un « small clearance » (piston B) pour terminer l'homogénéisation (**Figure 6B**). La solution homogène a été centrifugée à 500 g pendant 5 min à 4°C pour éliminer le tampon de lyse. Le culot a été re-suspendu dans 1 mL de PBS, puis filtré à travers un tamis cellulaire de 40 µm (Fisherbrand™). Un aliquot de 10 µL de solution filtrée a été prélevé pour le comptage, qui a été réalisé grâce à une coloration au Bleu Trypan 0,4% (Invitrogen™). 75 000 noyaux de chaque échantillon, comptés à l'aide d'une lame de comptage Kova, ont été collectés et centrifugés à 500 g pendant 5 min à 4°C. Après avoir éliminé le surnageant, les culots de noyaux ont été refroidis progressivement (-1°C/min) jusqu'à -80°C dans un portoir réfrigéré pour la congélation des cellules (Corning® CoolCell™) et stockés à -80°C jusqu'à leur utilisation.

Préparation des bibliothèques ATAC-seq

Les bibliothèques ont été préparées comme décrites ci-dessous, selon les étapes du protocole du kit Active Motif (ATAC-Seq Kit #53150) sans modifications. Le protocole est résumé **Figure 7A**

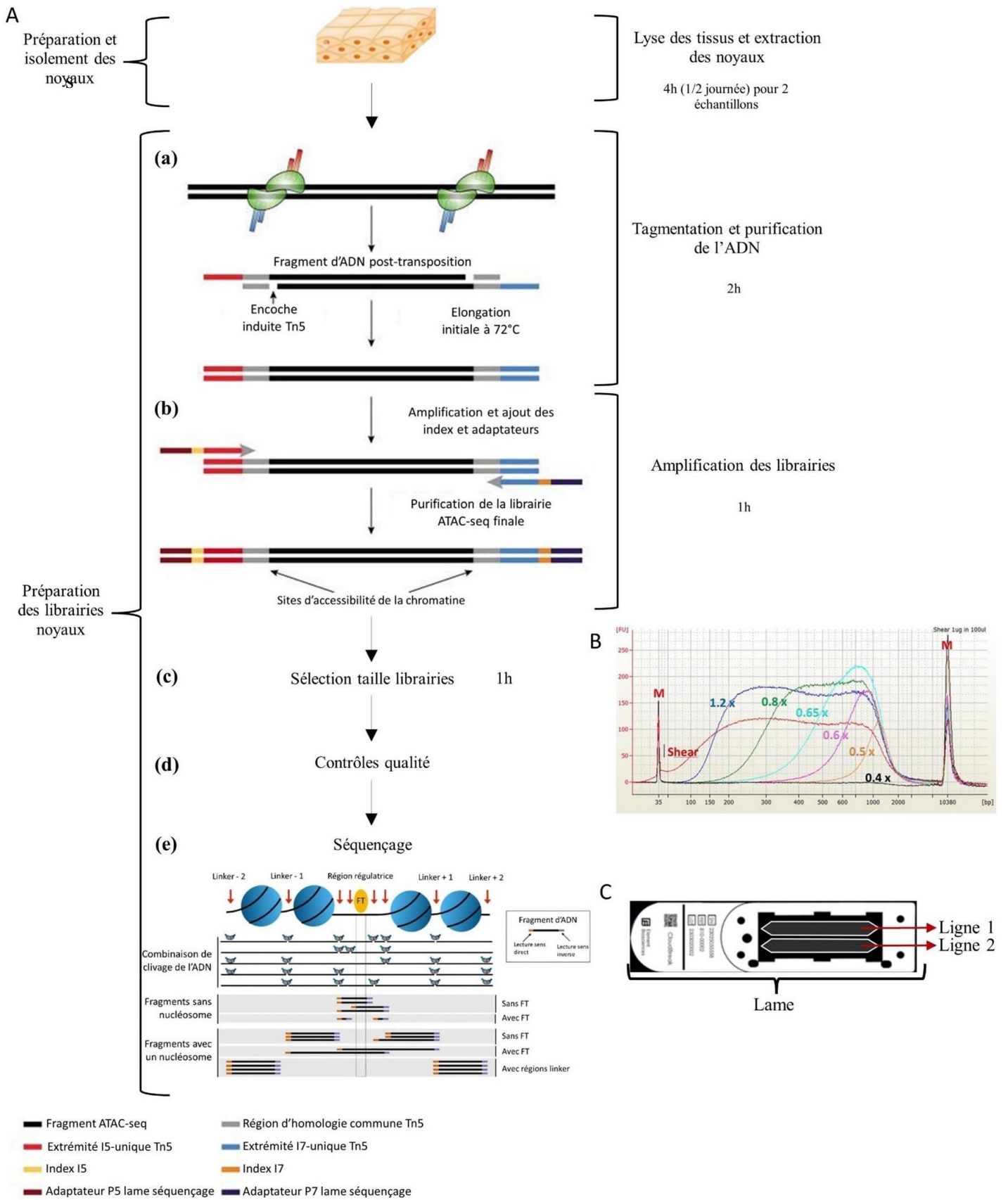


Figure 7 : Préparation des bibliothèques. A. Schéma des étapes pour la préparation des bibliothèques (modifié d'après Grandi et al. 2022). (e) Schéma des tailles de fragments attendus après séquençage (modifié d'après Zhijian et al. 2019) : 20 à 90 pb (fragments sans nucléosome), 146 pb (fragment avec un nucléosome). B. Graphique représentant les tailles de fragments retenues suivant le ratio billes/ADN utilisé pour la sélection de taille des bibliothèques. C. Représentation d'une lame AVITI composée de 2 lignes pour le séquençage.

(a) Le culot de chaque échantillon a été re-suspendu dans 50 μL de Tagmentation Master Mix (25 μL de 2X Tagmentation Buffer, 2 μL de 10X PBS, 0,5 μL de 1,0% Digitonin, 0,5 μL de 10% Tween 20, 10 μL de Transposase Tn5 et 12 μL H_2O). Les échantillons ont été incubés dans un thermomixer (Eppendorf ThermoMixer F2.0) à 37°C pendant 30 min à 800 rpm. Après la réaction de tagmentation, chaque échantillon a été transféré dans un nouveau tube, 250 μL de DNA Purification Binding Buffer et 5 μL d'acétate de sodium 3M ont été ajoutés. Suite à cela, chaque échantillon a été transféré dans une colonne de purification d'ADN. Les colonnes ont été centrifugées à 17 000 g pendant 1 minute. Un lavage des colonnes a été réalisé avec 750 μL de DNA Purification Wash Buffer et centrifugées à 17 000 g pendant 1 min. Les colonnes ont ensuite été séchées pendant 2 min à 17 000 g. L'ADN a été élué dans 35 μL de DNA Purification Elution Buffer.

(b) Une réaction d'amplification de 10 cycles est ensuite réalisée pour obtenir une quantité suffisante de matériel pour le séquençage, mais également pour l'incorporation des index servant à l'identification de chaque échantillon. Pour cela, 33,5 μL d'ADN purifié a été ajouté à 2,5 μL d'amorce index I7 (25 μM), 2,5 μL d'amorce index I5 (25 μM), 1 μL de dNTP, 10 μL de tampon 5X Q5 et 0,5 μL de polymérase Q5 (2 U/ μL). La PCR a été effectuée dans un thermocycleur 2720 Thermal Cycler (Applied Biosystems). Le programme utilisé est le suivant : 5 min à 72°C, 30 sec à 98°C et 10 cycles comprenant 10 sec à 98°C, 30 sec à 63°C et 1 min à 72°C.

(c) Par la suite, une étape de purification avec 60 μL de billes magnétique SPRI (Beckman coulter) a été réalisée pour chaque échantillon. Le ratio billes/ADN était de 1,2X afin de lier majoritairement les fragments ayant une taille supérieure à 100 pb (**Figure 7B**). Pour permettre la liaison des billes avec l'ADN, les tubes ont été incubés 5 min. Afin d'éliminer le surnageant, les échantillons ont été placés sur un portoir magnétique. Un double lavage a été réalisé avec 180 μL d'éthanol 80% pendant 30 sec, suivi d'un séchage de 5 min. Après avoir retiré les tubes

du portoir magnétique, les échantillons ont été purifiés dans 20 µL de DNA Purification Elution Buffer et incubés 5 min. Pour récupérer l'ADN élué, les tubes ont été de nouveau placés sur un portoir magnétique.

(d) Suite à cela, une quantification des librairies a été réalisée par dosage au Qubit 3 Fluorometer (Thermo Fisher Scientific) avec l'utilisation du Qubit™ dsDNA HS Assay Kits (#Q32851). La distribution de taille des librairies amplifiées par PCR a été analysée au Fragment Analyzer. Les échantillons ont ensuite été mélangés de manière à obtenir un mélange équimolaire qui a été séquencé.

Séquençage

Le séquençage du pool des 12 échantillons (2 génotypes * 2 échantillons * 3 tissus) a été réalisé sur la Plateforme GeT-PlaGe du centre Toulouse-Occitanie (<https://doi.org/10.1038/s41597-025-04769-4>) à l'aide de l'AVITI d'Element Biosciences. Le kit recommandait 40 millions de lectures par échantillons pour l'analyse de la chromatine ouverte et 100 millions de lectures par échantillon pour l'étude du positionnement des nucléosomes. Une lame AVITI, composé de 2 lignes (**Figure 7C**), permet d'obtenir 800 millions de lectures. Il est donc attendu d'obtenir environ 66 millions de lectures par échantillon. Les lectures ont été séquencées par paires (« paired-end ») 2x75 pb.

Traitements des données ATAC-seq

Le cluster Genobioinfo, administré par la plateforme Genotoul-Bioinfo du centre Toulouse-Occitanie, a été utilisé et les données ont été traitées avec le pipeline nfcore-atacseq en utilisant les paramètres par défaut. Pour la sélection des régions de chromatine ouverte (peak calling), le paramètre *macs_fdr* a été fixé à 0,05. Ce paramètre sert à limiter les faux positifs : un *macs_fdr* faible sera plus stringent. Un fichier IGV (Integrative Genomics Viewer, version 2.19.4) a été généré. Les fichiers bigwig de tous les échantillons ont été chargés sur les pistes.

HPRT

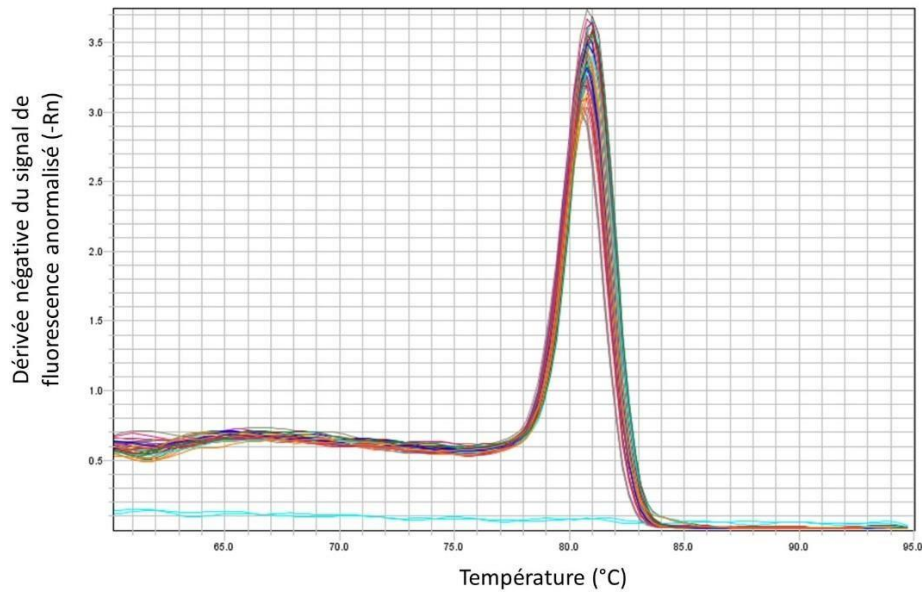
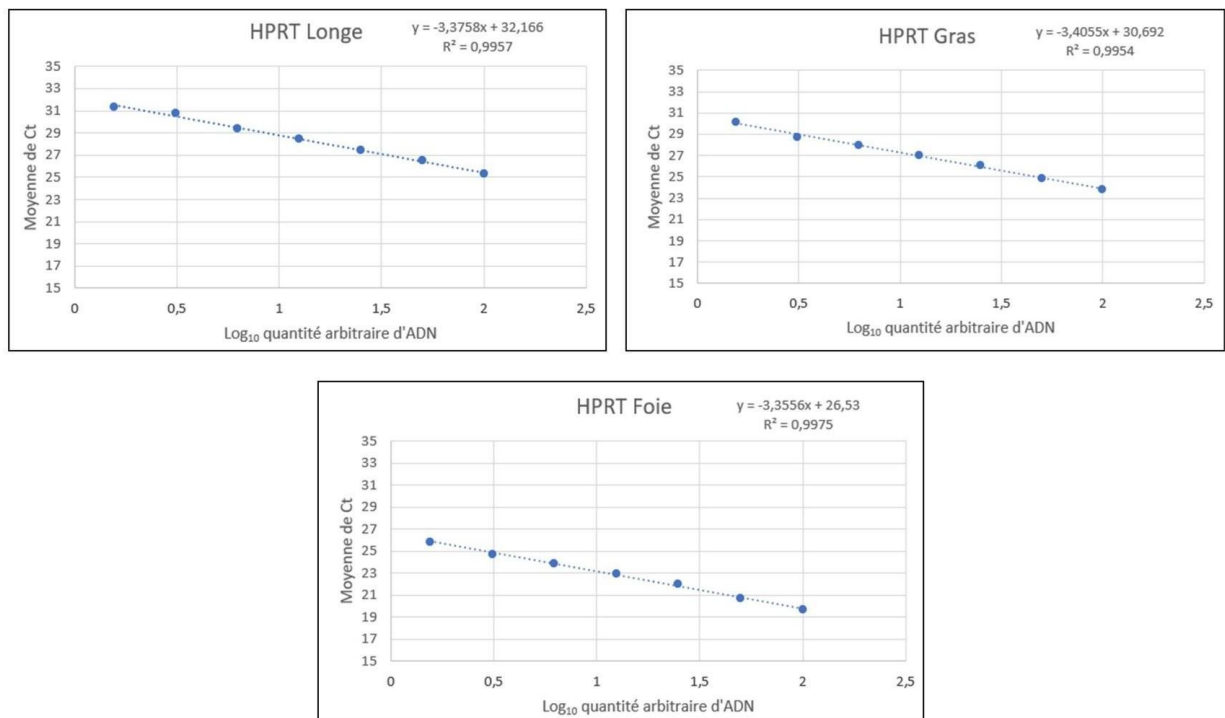


Figure 8 : Courbe de dissociation pour le gène *HPRT*. Mesure de la variation de fluorescence en fonction de la température. Le tracé bleu correspond au contrôle négatif. Un pic unique indique la présence d'un seul amplicon spécifique. Profil observé pour tous les gènes de références dans les 3 tissu.

A



B

	HPRT			RPL32			TBP1			B-Act			TOP2B		
	Longe	Gras	Foie	Longe	Gras	Foie	Longe	Gras	Foie	Longe	Gras	Foie	Longe	Gras	Foie
Pente	-3,38	-3,41	-3,36	-3,46	-3,37	-3,31	-3,99	-3,25	-3,60	-3,51	-3,79	-3,48	-3,04	-2,74	-2,70
R ²	1,00	1,00	1,00	1,00	1,00	1,00	0,99	0,99	0,99	0,99	0,99	0,98	0,99	1,00	0,98
Eff	1,98	1,97	1,99	1,94	1,98	2,01	1,78	2,03	1,90	1,93	1,84	1,94	2,13	2,31	2,34
%Eff	97,80	96,63	98,61	94,36	98,18	100,58	78,16	103,01	89,69	92,74	83,61	93,98	113,10	131,46	134,49

Figure 9 : Mesure l'efficacité des couples d'amorces. **A.** Droite de régression pour le gène *HPRT* dans les 3 tissus. En abscisse est représenté le \log_{10} d'une quantité arbitraire d'ADN et en ordonnée la moyenne de Ct entre chaque duplicat. **B.** Tableau regroupant les efficacités de tous les gènes de références pour les 3 tissus, calculé avec la formule suivante : $E = 10^{-1/\text{pente}}$.

L'option « auto-scale » a été appliquée pour régler l'intensité de signal

Résultats

Partie 1 : Etude de l'expression des 6 gènes d'intérêt

Dans un premier temps, des analyses d'expression ont été réalisées afin de déterminer s'il existe un différentiel d'expression parmi les six gènes d'intérêt, pour les 10 individus Rec/MS et les 10 individus MS/MS.

1.1. Détermination de l'efficacité des couples d'amorces

Déterminer l'efficacité des couples d'amorces pour chacun des gènes est nécessaire pour optimiser les conditions expérimentales et obtenir des résultats fiables et reproductibles. L'efficacité des amorces des gènes d'intérêt avait déjà été validée, donc seule l'efficacité des couples d'amorces de 5 gènes de référence a été testée : *HPRT*, *TBPI*, *RPL32*, *ACTINEB* et *TOP2B*. L'efficacité permet de vérifier que les couples d'amorces utilisés sont fonctionnels et que l'amplification est spécifique via la présence d'un seul produit d'amplification. Pour cela, les courbes de dissociation sont d'abord validées pour l'ensemble des gènes. Tous les gènes de références montrent une amplification spécifique comme le gène *HPRT* (**Figure 8**). L'efficacité des amorces peut s'exprimer en pourcentage. Elle est optimale lorsque la quantité d'ADN double à chaque cycle. Néanmoins, une efficacité supérieure à 80% est considérée comme acceptable. Une droite de régression est tracée pour déterminer l'efficacité. Un exemple de droites de régression est représenté **Figure 9A** pour le gène *HPRT* pour les 3 tissus. Les efficacités des couples d'amorces pour les gènes testés sont regroupées dans la **Figure 9B**. Les efficacités de tous les gènes se situent entre 80 et 100% pour les 3 tissus à l'exception de *TOP2B*. En effet, le gène *TOP2B* présente des efficacités supérieures à 100%, indiquant une amplification plus importante que le doublement de l'amplicon. Ce gène a donc été mis de côté et les autres gènes ont été conservés.

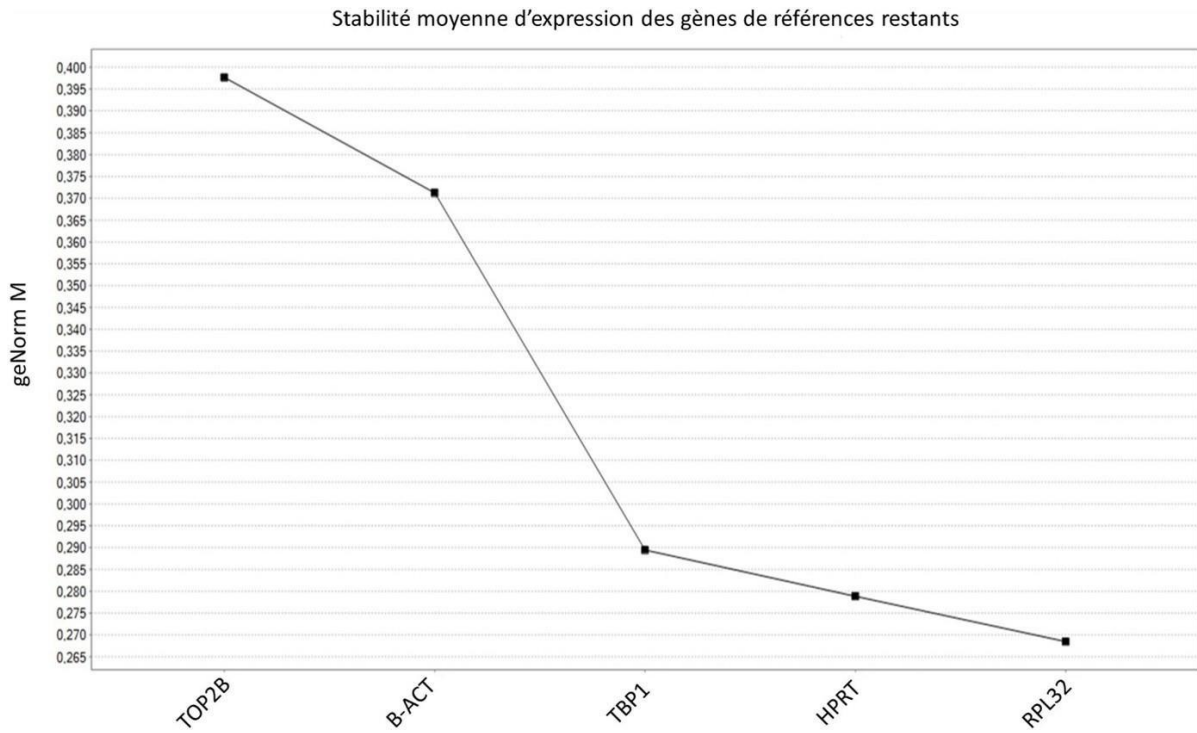


Figure 10 : Analyse de la stabilité des gènes de référence réalisée avec geNorm. Les valeurs M représentent la stabilité d'expression de chaque gène candidat. Les gènes sont classés par ordre décroissant de stabilité.

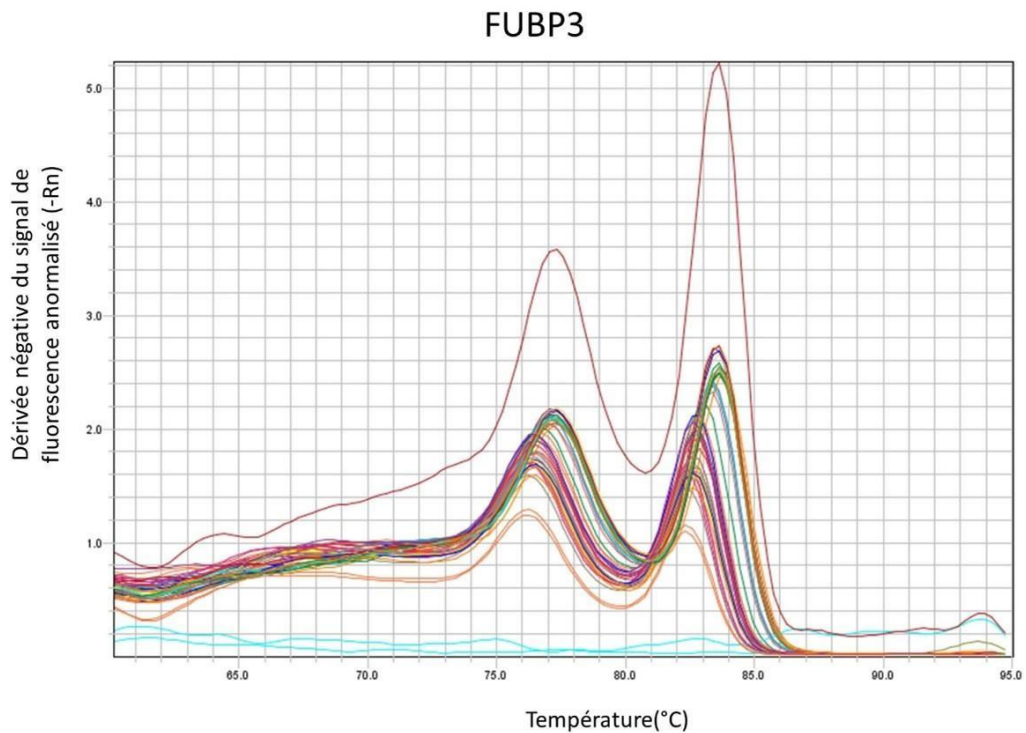


Figure 11 : Courbe de dissociation pour le gène *FUBP3*. Mesure de la variation de fluorescence en fonction de la température. Le tracé bleu correspond au contrôle négatif. Deux pics indiquent la présence de produits non spécifiques. Ce profil est observé pour ce gène dans les 3 tissus.

1.2. Sélection des gènes de références

L'expression d'un gène d'intérêt est mesurée relativement à l'expression d'un (ou plusieurs) gène(s) de référence. La stabilité des 4 gènes de référence a été testée sur les 20 échantillons de longe. La stabilité peut être variable en fonction des tissus. Cependant les mesures n'ont pas été réalisées sur le tissu adipeux et le foie, car elles avaient déjà été faites préalablement. La stabilité des gènes de référence est évaluée avec la méthode geNorm qui compare chaque gène aux autres en calculant une valeur appelée « valeur M ». Plus la valeur M est basse, plus le gène est stable. À chaque étape, le gène le moins stable est retiré, et le calcul recommence jusqu'à ce qu'il ne reste que deux gènes. Un gène est considéré comme stable si sa valeur M est inférieure à 0,5.

Les résultats représentés **Figure 10** classent *RPL32* comme le plus stable (environ 0,267), suivi de *HPRT* (environ 0,278) et *TBPI* (environ 0,289). L'*ACTINEB* et *TOP2B* ont les moins bons scores avec des valeurs M d'environ 0,372 et 0,397 respectivement. Les 3 gènes de références les plus stables sont *RPL32*, *HPRT* et *TBPI*. Cependant des qPCR avaient déjà été réalisées préalablement. Le choix a été de retenir les mêmes gènes de références, à savoir *HPRT* et *TBPI*.

1.3. Analyse de l'expression

Parmi les 3 gènes présents dans l'intervalle (candidats positionnels) et les 3 gènes en dehors, *GPR107* et *FNBPI* sont les 2 candidats fonctionnels. L'allèle récessif MS est favorable pour la croissance et le gras. Une surexpression de *FNBPI* dans la longe et de *GPR107* dans le tissu adipeux sont attendues, avec un effet plus important chez les MS/MS que Rec/MS. En effet, les MS/MS possèdent 2 allèles favorables et récessifs permettant l'expression du phénotype. L'expression relative des 6 gènes d'intérêt est calculée par la méthode des 2 delta CT sur les 10 échantillons Rec/MS et les 10 MS/MS pour les 3 tissus. Pour les 3 tissus, *FUBP3* montre une amplification non spécifique (**Figure 11**). Ce gène est donc exclu pour la suite de l'analyse. L'expression différentielle est validée par un test de Student bilatéral. Pour la longe et le tissu

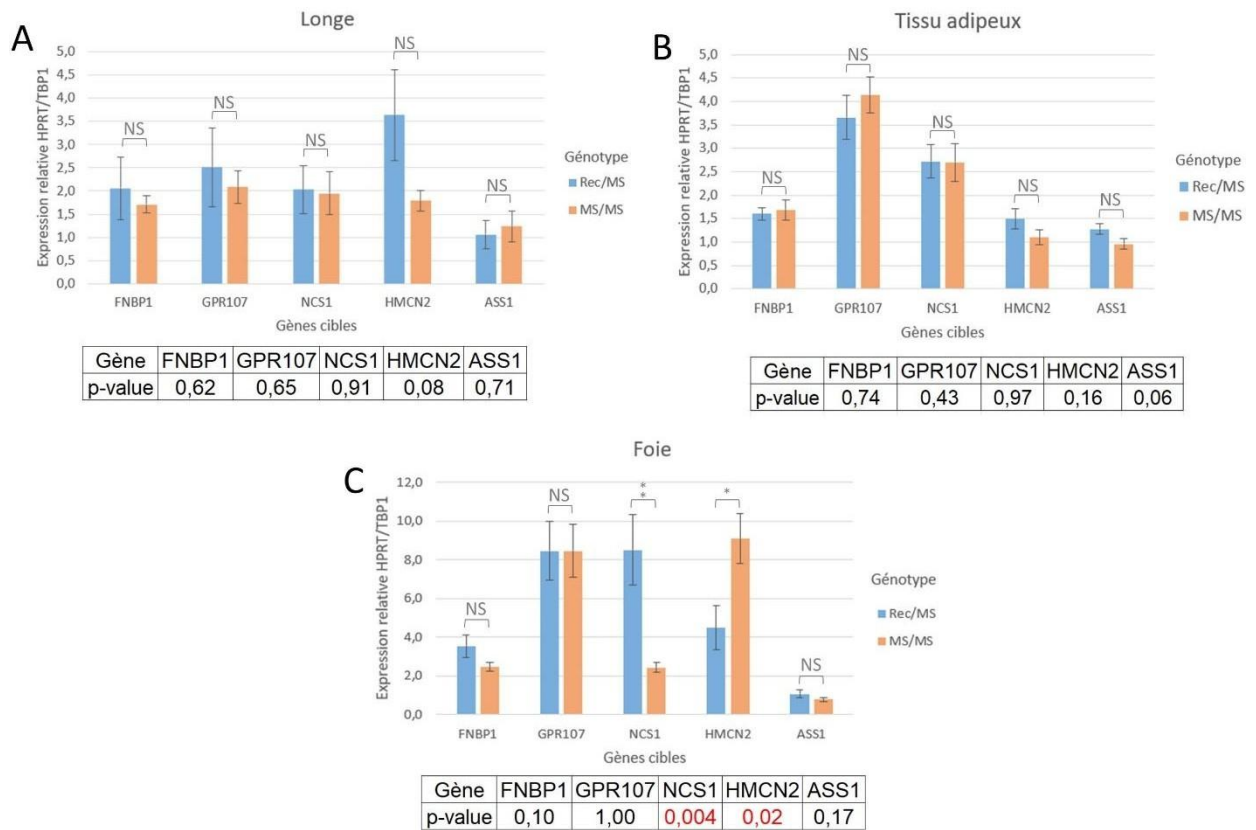


Figure 12 : Expression relative des 6 gènes d'intérêt dans les 3 tissus. Expression par génotype des 5 gènes relatif à la moyenne géométrique des gènes *HPRT* et *TBP1* pour chaque tissu : longe (A), tissu adipeux (B) et foie (C). Les p-values, représentées dans les tableaux, sont calculées par un test de Student ; NS = non significatif, * = $p < 0.05$ et ** = p-value < 0.01 .

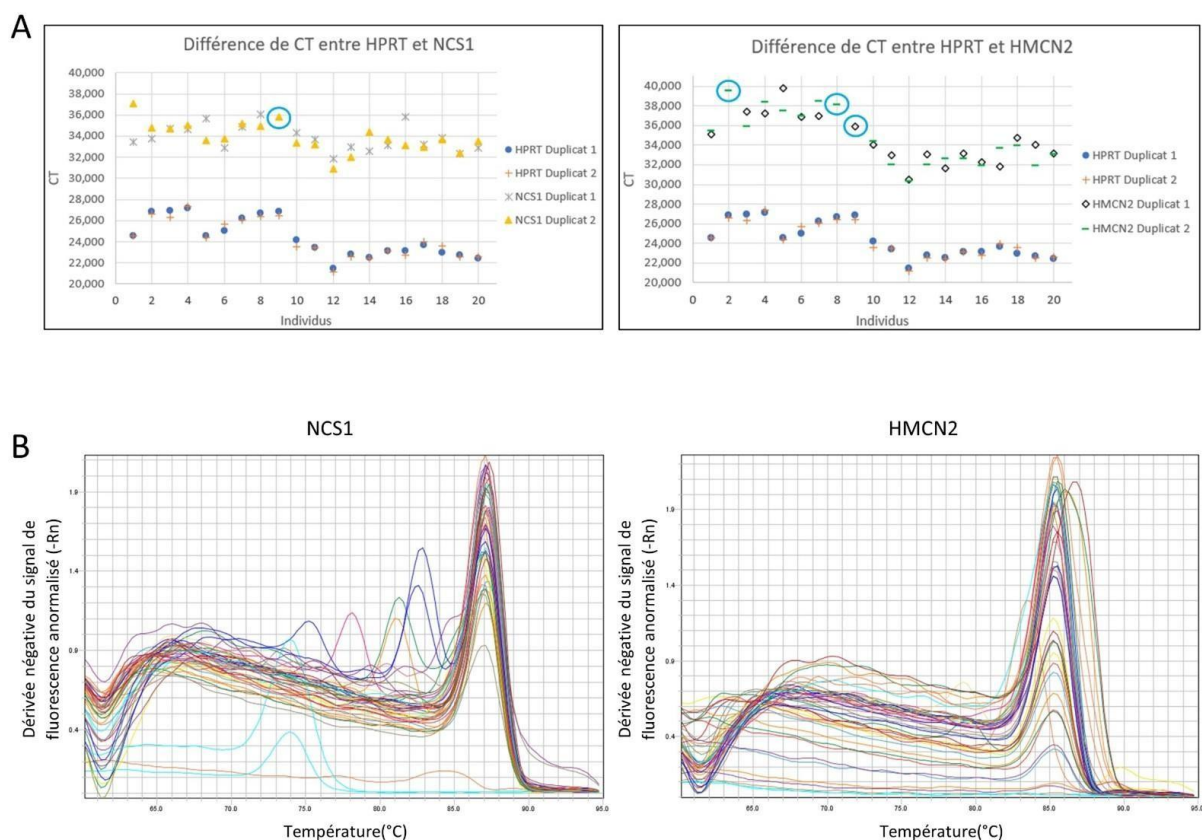


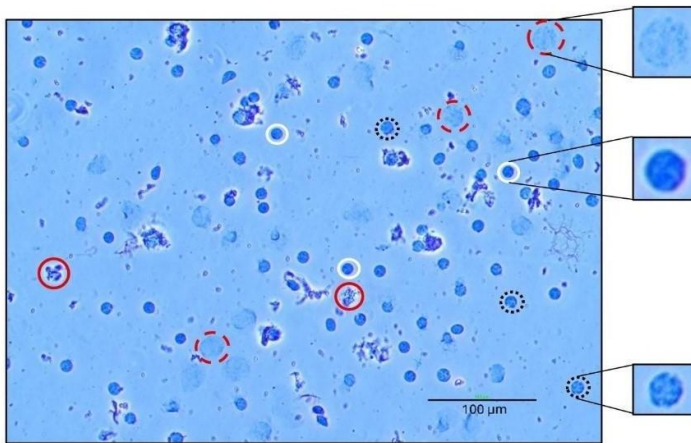
Figure 13 : Contrôle des données pour le foie. A. Mesure de la différence de Ct pour chaque duplicat entre *HPRT* et *NCS1* (à gauche) ou *HPRT* et *HMCN2* (à droite). Les cercles bleus indiquent la présence d'un seul duplicat. B. Courbe de fusion pour les gènes *NCS1* et *HMCN2*.

adipeux, aucun des 5 gènes ne présente un différentiel d'expression significatif bien qu'une tendance soit observée pour les gènes *HMCN2* et *ASS1* respectivement dans la longe et le tissu adipeux (**Figure 12A et B**). Pour le foie, 2 gènes montrent un différentiel d'expression : *NCSI* (p-value = 0,004) et *HMCN2* (p-value = 0,015) (**Figure 12C**). Cependant, les résultats très différents entre tissus et parfois non concordants, comme pour *HMCN2* qui montre des effets opposés dans la longe et le foie, nous ont amené à évaluer la qualité des résultats, bien que les conditions expérimentales pour l'amplification de ces gènes avaient été validées préalablement. Chaque échantillon a été traité en duplicat. Les valeurs de cycles seuil (Ct) sont comparées entre duplicats afin d'identifier d'éventuels écarts de Ct. Un écart inférieur à 1 Ct est toléré. Les courbes de dissociation sont également vérifiées. Pour les gènes *NCSI* et *HMCN2*, la qualité des données est validée pour la longe et le tissu adipeux mais pas pour le foie. En effet, ces deux gènes présentent des Ct très tardifs, aux alentours de 34 Ct et 36 Ct respectivement. Les écarts entre duplicats sont plus importants et parfois un duplicat est manquant (**Figure 13A**). En principe, les Ct sont compris entre 22 et 30. Au-delà, les données sont moins sensibles et donc moins fiables. Cela est également visible sur les courbes de dissociation (**Figure 13B**) dont les pics de fusion varient, reflétant une différence de quantité des produits amplifiés. La qualité des données n'étant pas validée pour ce tissu et ces deux gènes, les résultats restent donc à confirmer.

Partie 2 : Etude de l'accessibilité de la chromatine

Une analyse ATAC-seq a été effectuée afin d'identifier les régions de chromatine accessible, c'est-à-dire, les régions qui ne sont pas compactées et donc accessible aux protéines régulatrices. Les expériences d'ATAC-seq nécessitent des noyaux ayant conservé une membrane nucléaire intacte afin de préserver l'état chromatinien. Cette analyse permettra également de confirmer les résultats obtenus par qPCR concernant particulièrement les gènes localisés dans la région d'intérêt du chromosome 1.

A



B

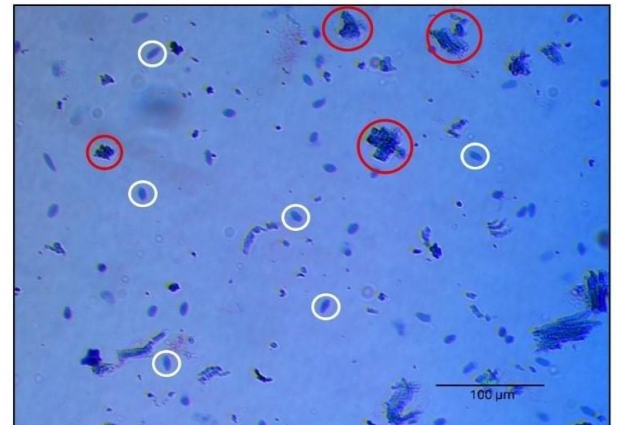
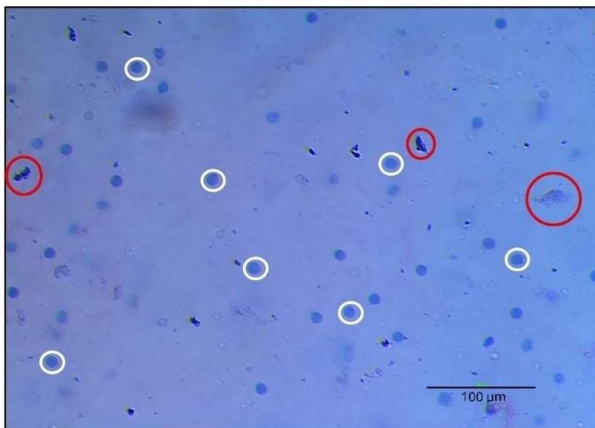


Figure 14 : Observation au microscope des noyaux avant et après optimisation. **A.** Avant optimisation : culot nucléaire pur et observé à l'objectif 20X pour le foie (gauche) ; culot nucléaire pur et observé à l'objectif 10X pour la longe (droite). De nombreux débris sont observés (cercles rouges) ainsi que quelques noyaux avec (cercles blancs) et sans membrane (cercles rouges en pointillés). La membrane de certains noyaux semble partiellement cassée (cercles pointillés noirs) **B.** Après optimisation : culot nucléaire dilué au 1/10 et observé à l'objectif 20X pour le foie (gauche) ; culot nucléaire dilué au 1/2 et observé à l'objectif 20X pour la longe (droite). Beaucoup de débris et beaucoup plus de noyaux sont observés.

2.1. Optimisation du protocole Active Motif

Une première expérience a été réalisée en suivant le protocole Active Motif sur des échantillons de foie. Les premiers résultats obtenus ne sont pas totalement satisfaisants : beaucoup de débris cellulaires sont observés. Des noyaux avec une membrane nucléaire sont présents, mais, certaines semblent altérées. En effet, comme le montre la (**Figure 14A, gauche**), certains noyaux semblent dépourvus de leur membrane ; ils sont reconnaissables par leur taille et une coloration plus claire, en raison de la diffusion du Bleu Trypan. Le protocole a aussi été réalisé sur des échantillons de longe. Cependant, le tissu s'est mal dissocié et un grand nombre de débris est visible (**Figure 14B, droite**). Une optimisation du protocole a donc été nécessaire. Pour ce faire, le protocole d'Actif Motif a été comparé au protocole du kit Diagenode (Tissue Nuclei Extraction for ATAC-seq #C01080003). Les étapes du kit Diagenode, jugées pertinentes pour améliorer le protocole, ont été appliquées au protocole Active Motif. Après le découpage des tissus, l'étape de lavage a été retirée. Elle n'est pas présente dans le protocole du kit Diagenode et ne semble pas améliorer l'aspect des noyaux. Les principales modifications ont porté sur le temps de lyse et la dissociation. En effet, la spécificité de chaque tissu entraîne des conditions différentes :

- Pour l'étape de lyse, une durée de 3 min et 10 min ont été nécessaires pour les échantillons de foie et de tissu adipeux. Pour la longe, une étape de lyse à de l'acutase a été ajoutée avant les 10 min de tampon de lyse du kit. Les échantillons ont été traités avec 1 mL d'acutase pendant 15 min à 37°C ;
- Pour l'étape de dissociation avec le dounce, les mêmes modifications ont été appliquées pour les 3 tissus : le piston large clearance (piston A) a été ajouté avant l'utilisation du piston small clearance (piston B) et le nombre d'allers-retours suivant les pistons est déterminé en suivant le kit Diagenode : 10 x large clearance et 20 x small clearance.

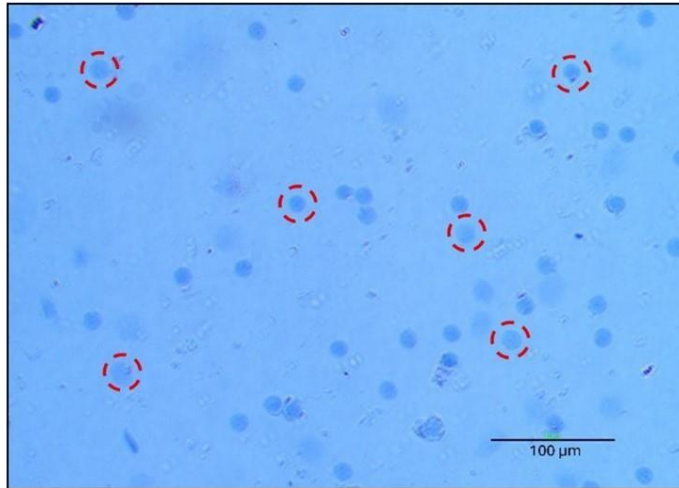


Figure 15 : Observation du culot nucléaire après gradient d'iodixanol. Culot nucléaire dilué au 1/10 et observé à l'objectif 20X pour le foie. Des noyaux dépourvus de membrane nucléaire sont observés (cercles rouges en pointillés).

Par la suite, une deuxième filtration a été ajoutée ainsi qu'une centrifugation finale, comme indiqué dans le protocole Diagenode. Les échantillons de foie ont été dilués au 1/10^{ème} et les échantillons de tissu adipeux et de longe au demi. Une coloration au Bleu Trypan a ensuite été réalisée. Les résultats obtenus pour les 3 tissus sont très satisfaisants. Pour les échantillons de foie, des noyaux ronds et possédant leur membranes nucléaires intactes sont observés. Des débris cellulaires et des noyaux sans membrane nucléaire sont présents mais en très faible proportion (**Figure 14B, gauche**). Les noyaux des échantillons de tissu adipeux possèdent leur membrane nucléaire mais ils sont présents en quantité plus faible. Pour les échantillons de longe, les noyaux sont caractérisés par une forme ovale, caractéristique du tissu musculaire squelettique. Des débris cellulaires sont aussi observés pour tous ces échantillons (**Figure 14B, droite**). La qualité des noyaux étant validée, ils ont été comptés manuellement à l'aide d'une lame de comptage Kova.

Après l'extraction des noyaux, le kit Diagenode inclut une étape de purification des noyaux par gradient d'iodixanol. Cela permet d'isoler uniquement les noyaux, sans débris cellulaire. Cette étape est appliquée sur les noyaux obtenus précédemment. Aucun débris n'est détecté lors de l'observation au microscope. Seuls des noyaux sont prélevés. Toutefois, ces derniers ne possèdent plus leur membrane nucléaire (**Figure 15**) suggérant que l'état chromatinien a pu être affecté. Ces résultats peuvent s'expliquer par le fait que le laboratoire ne possède pas le type de centrifugeuse recommandé par le protocole. Ne pouvant donc pas effectuer l'étape de gradient dans les conditions optimales recommandées par le kit Diagenode (Tissue Nuclei Extraction for ATAC-seq #C01080003) cette étape n'a pas été réalisée. Les noyaux étant de bonne qualité avant l'étape de purification par gradient d'iodixanol, l'étape de construction des bibliothèques a été réalisée à partir des échantillons précédents, malgré les quelques débris cellulaires restants.

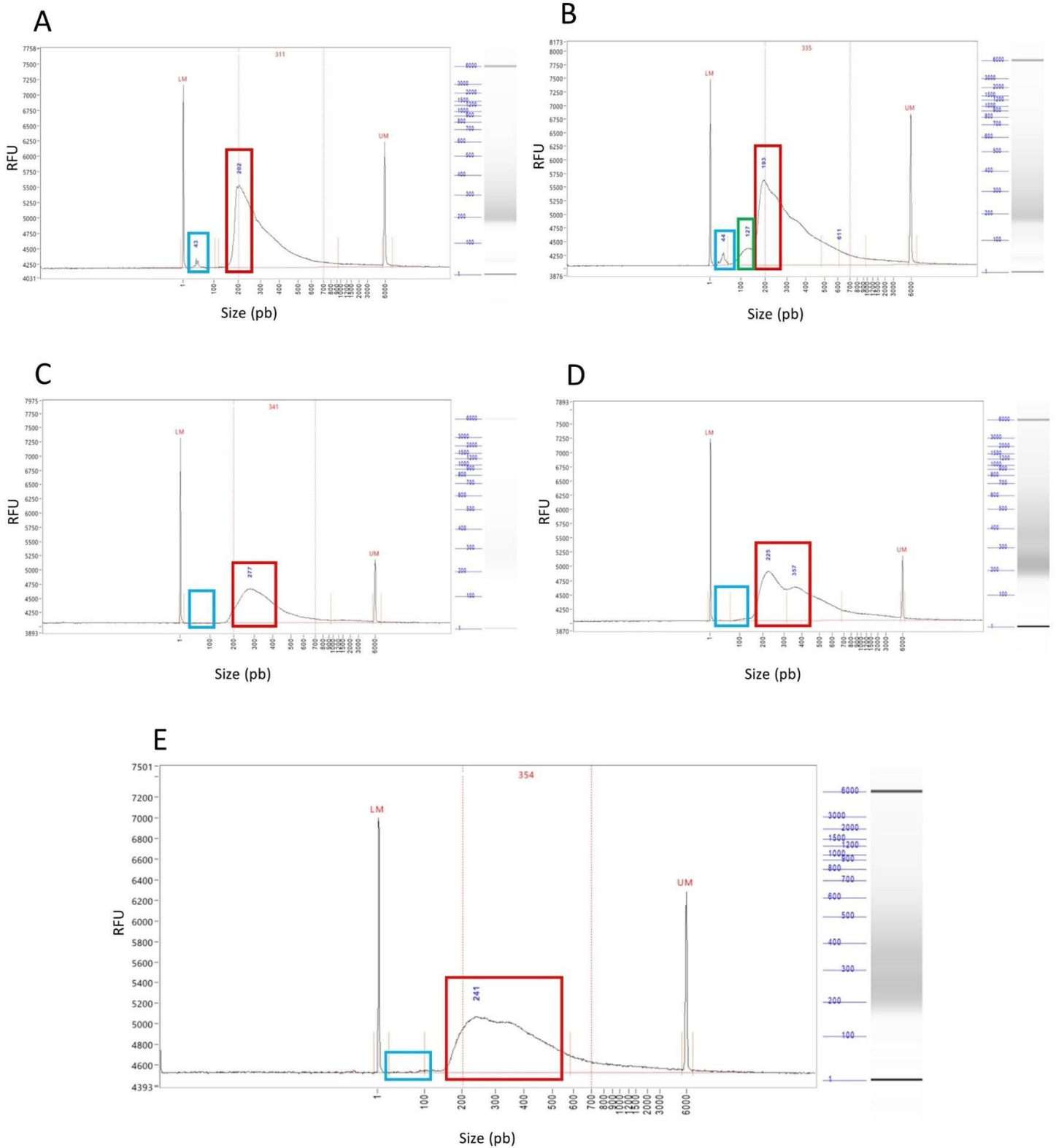
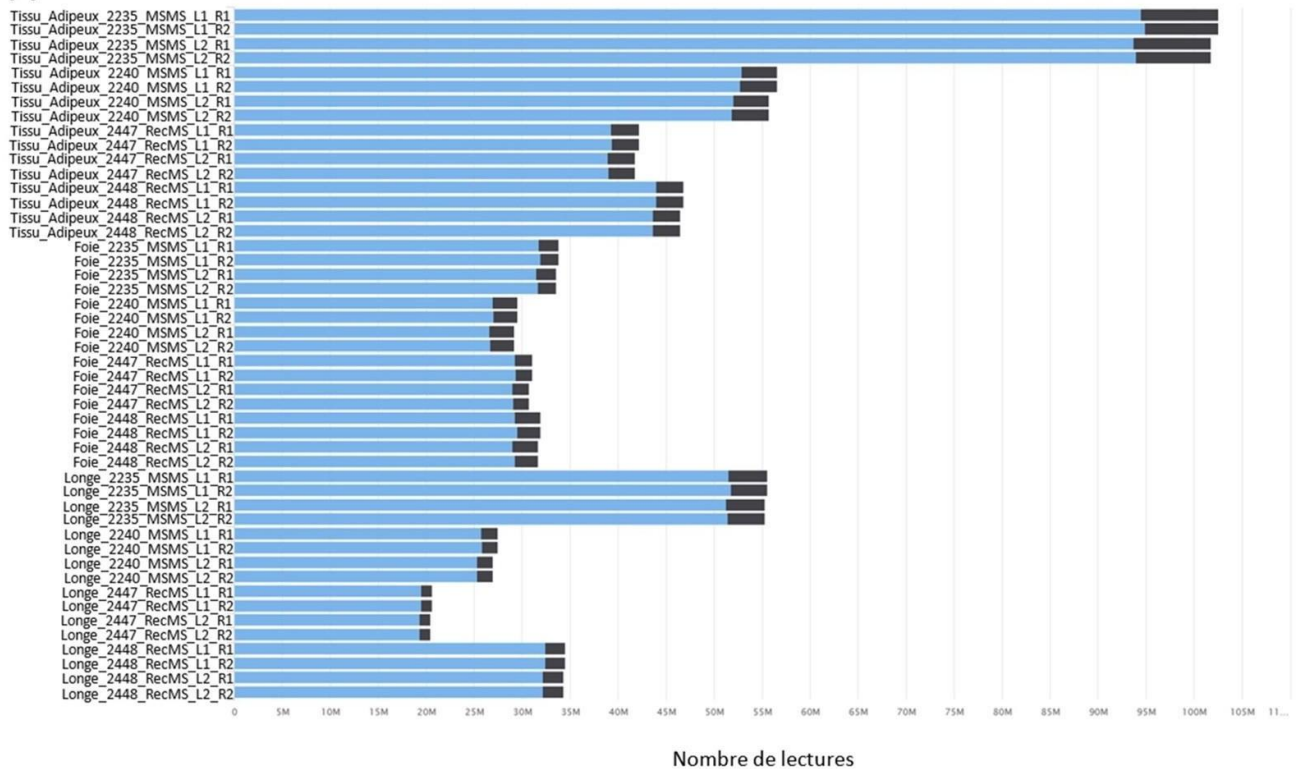


Figure 16 : Profils des bibliothèques obtenues au Fragment Analyzer. A. Profil des bibliothèques d'un échantillon de foie après une purification 1,2X. **B.** Profil des bibliothèques d'un échantillon de gras après une purification 1,2X. **C.** Profil des bibliothèques d'un échantillon de foie après deux purifications : 1,2X et 0,9X. **D.** Profil des bibliothèques d'un échantillon de gras après deux purifications : 1,2X et 0,9X. **E.** Profil des bibliothèques du pool d'échantillon après deux purifications : 1,2X et 0,9X. La taille des bibliothèques est encadrée en rouge, la présence ou l'absence de dimères d'amorces est encadrée en bleu et l'encadré vert montre la présence d'un épaulement non souhaité.

2.2. Contrôle qualité des librairies

Les librairies ont été construites selon les recommandations du kit Active Motif (ATAC-Seq Kit #53150) avant d'en évaluer la qualité. Les fragments correspondant aux séquences sans nucléosome ont une taille variable de 20 à 90 pb. Pour les fragments correspondant à 1 nucléosome la taille est de 146 pb. La taille d'un fragment correspondant à 2 nucléosomes est de 292 pb et ainsi de suite (**Figure 7A, (e)**). De plus, lors de la construction de la librairie, des adaptateurs et des index sont ajoutés de chaque côté des fragments pendant la réaction d'amplification. Cela rajoute une centaine de pb à la taille du fragment. La taille des librairies doit varier entre 200 et 800 pb. Les profils obtenus sur le Fragment Analyzer après une première purification, montrent une distribution de taille entre 200 et 800 pb pour tous les échantillons. Un pic d'intensité plus élevée est observé de 150 à 250 pb comme montré sur les **Figures 16A et B** qui représentent un échantillon de foie et un échantillon de tissu adipeux. Cependant, un pic d'environ 40 pb, correspondant à des dimères d'amorces, est observé pour tous les échantillons et un épaulement à 120 pb est observé pour les échantillons de tissu adipeux (**Figure 16B**). C'est pourquoi une seconde purification avec les billes SPRI a été réalisée en utilisant un ratio de 0,9X billes/ADN pour éliminer les fragments inférieurs à 150 pb (**Figure 7B**). Le profil des librairies a été corrigée après cette seconde purification ; la taille des fragments varie entre 200 pb à 700 pb. Pour le foie, un unique pic est détecté de 200 à 450 pb avec un maximum de fragments présentant une taille de 202 pb suggérant majoritairement des fragments sans nucléosome (**Figure 16C**). Pour le tissu adipeux, deux pics sont observés : à 200 pb et 350 pb pouvant suggérer un enrichissement en séquences sans nucléosome et avec un nucléosome (**Figure 16D**). Concernant la longe, 2 échantillons présentent un profil similaire aux échantillons de foie et les 2 autres un profil similaire aux échantillons de tissu adipeux. Ensuite, les 12 échantillons sont mélangés de façon équimolaire et ce mélange est à nouveau passé au Fragment Analyzer. Le profil obtenu montre une répartition des tailles de 200 à 700

A



B

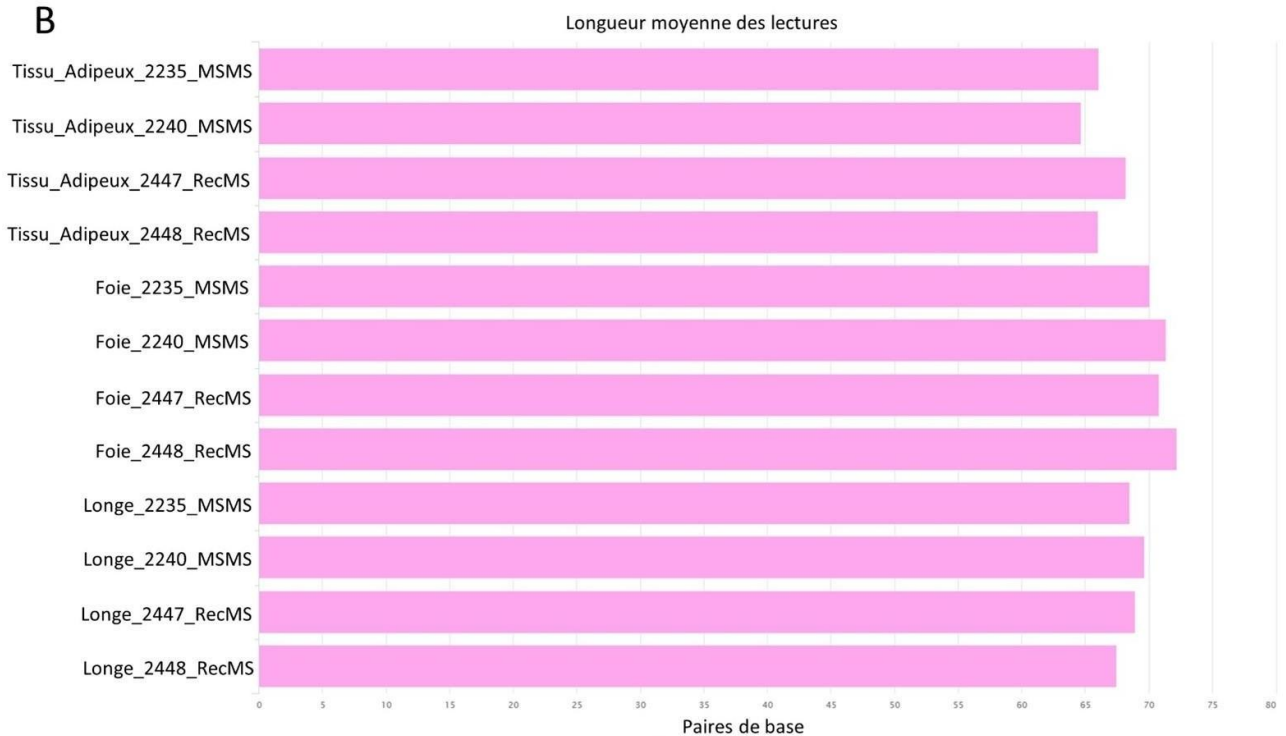


Figure 17 : Qualité des lectures de séquençage selon les échantillons et les tissus. A. Nombre des lectures pour chaque échantillons en millions (FastQC). Les échantillons sont nommés comme suit : tissu_individu_génotype_LigneDeSéquençage_lectures, bleu : lecture unique et noir : lecture dupliquée. **B.** Longueur des lectures pour chaque échantillon en paire de base (Picard Tools).

pb. Un pic large allant de 200 pb à 500 pb et une absence de dimères d'amorces (**Figure 16E**). Ces profils de librairies ont été validées pour poursuivre l'étape de séquençage.

2.3. Contrôles qualité des données ATAC-seq

Les 12 échantillons ont été séquencés en mode « paired-end » sur le séquenceur AVITI pour générer 400 millions de données par ligne. A l'issue du run, différents contrôles qualité ont été effectués sur les données obtenues, permettant de vérifier la qualité des données de séquençage et garantir la fiabilité des analyses bio-informatiques.

- Nombre de lecture par échantillon

Après le séquençage, 66 millions de lectures sont attendues pour chaque échantillon. Le nombre de lectures obtenu varie de 20 millions à 100 millions, suggérant une disparité entre les échantillons. Un échantillon est notamment surreprésenté, probablement due à des incertitudes de dosage en amont du mélange des échantillons (**Figure 17A**) ;

- Longueur des lectures

Le séquençage ayant été réalisé en 2x75 pb, la longueur des lectures attendue est de 75 pb. La taille des lectures varie entre 65 et 72 pb (**Figure 17B**) ;

- Alignement des lectures

Les lectures filtrées ont été alignées sur le génome de référence du porc (Sscr11.1). La **Figure 18A** montre que la proportion de lectures alignées est corrélée à la taille des chromosomes de référence. Les résultats montrent également peu de lectures alignées sur le chromosome mitochondrial, signifiant que la purification des noyaux a été efficace ;

- Taille des inserts

Les inserts correspondent aux fragments d'ADN génomique coupés par la transposase après avoir éliminé les adaptateurs et les index. Cette taille reflète donc directement l'accessibilité de la chromatine dans le noyau. Pour les 12 échantillons, 2 types de profils sont observés dans les

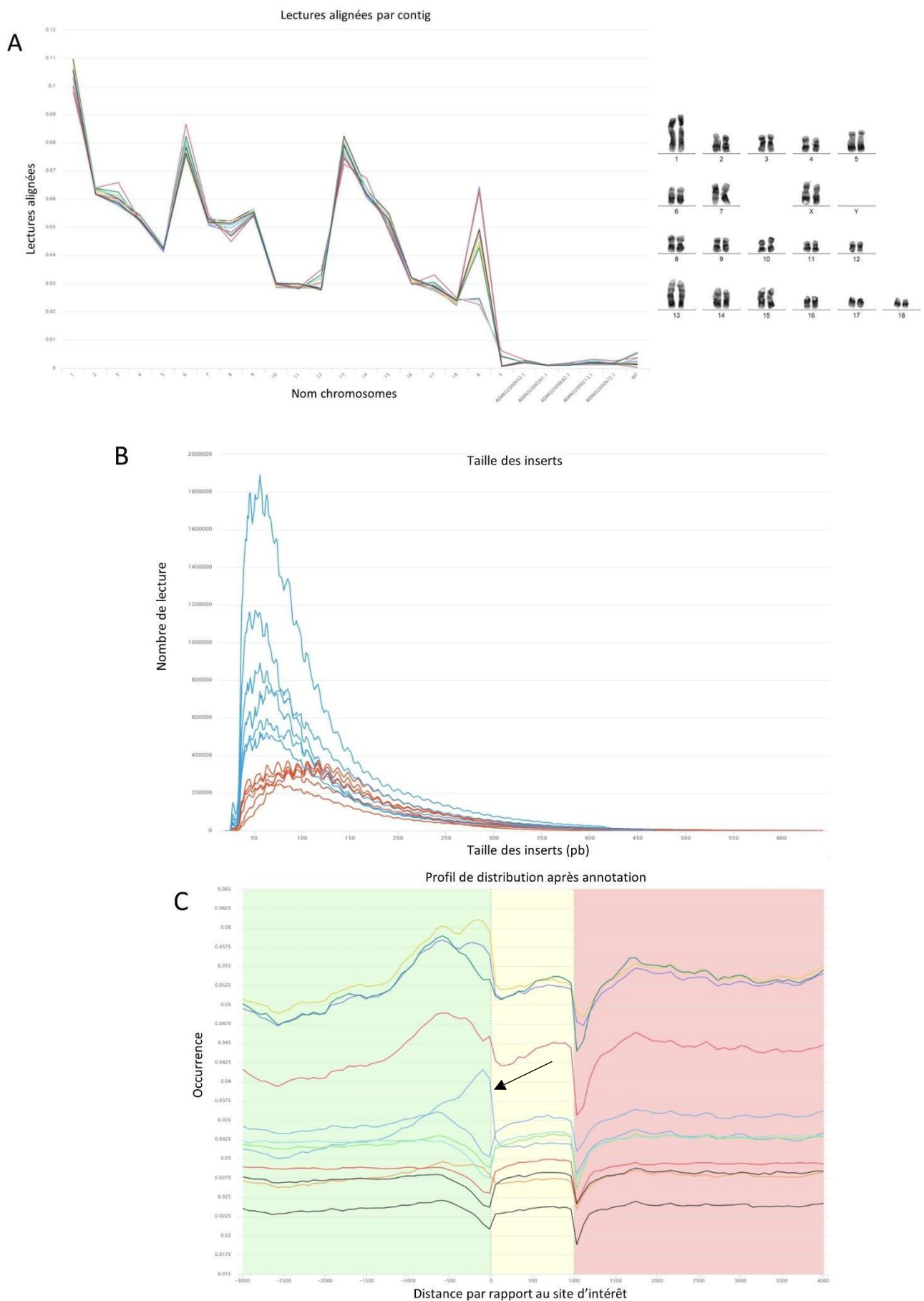


Figure 18 : Analyse des lectures de séquençage pour chaque échantillon et chaque tissus. **A.** Alignement des lectures pour chaque échantillon sur le génome (Sscr11.1) et sur le chromosome mitochondrial (Samtools) à gauche. Représentation d'un caryotype porcine à droite. **B.** Distribution des tailles d'inserts de chaque échantillon en paire de base. Bleu : profil avec 1 pic à 50/60 pb et rouge : profil avec un pic allant de 50 à 200 pb (Picard Tools). **C.** Profil de distribution autour des TSS. La zone verte correspond à la région promotrice, en amont du TSS. La zone jaune représente le corps du gène qui commence au TSS et fini au site de terminaison de la transcription (TES). La zone rouge correspond à la région en aval du gène (Deeptools). Flèche noire : échantillon de tissu adipeux qui semble montrer un enrichissement.

résultats obtenus. Le premier profil indique qu'une majorité d'inserts à une taille aux alentours de 50/60 pb, correspondant à des séquences sans nucléosome (**Figure 18B**). Les 4 échantillons de tissu adipeux et les 2 de longe, qui possèdent ce profil, présentent également le même profil de librairie avec 2 pics après la seconde purification (**Figure 16D**). Le second profil montre une taille d'inserts variant de 50 à 200 pb (**Figure 18B**). Il est composé des 4 échantillons de foie et des 2 autres échantillons de longe ; les librairies de ces échantillons présentaient une majorité de fragments à 277 pb après purification (**Figure 16C**) ;

- Profil de distribution après annotation

L'ensemble des gènes a été annoté sur le génome porcin (Sscr11.1) pour visualiser l'alignement des lectures le long des gènes en analysant la partie en amont du sites d'initiation de la transcription (TSS) et la partie en aval. Malheureusement, les résultats ne montrent pas d'enrichissement autour des TSS pour 6 échantillons (4 longues + 2 tissus adipeux) (**Figure 18C**). Les 6 autres échantillons (4 foies + 2 tissus adipeux) semblent montrer un léger enrichissement au niveau des TSS, surtout pour l'un des échantillons de tissu adipeux (indiqué par une flèche noire).

Enfin, une représentation IGV (Integrative Genomics Viewer) montre les résultats d'analyse du séquençage pour les 12 échantillons et focalisée sur le gène candidat positionnel *ASS1* localisé dans la région génomique du QTL (**Figure 19**). Toutes les pistes semblent montrer un alignement homogène des lectures sur la région qui inclut les zones en amont et aval du site d'initiation de la transcription du gène *ASS1*. Ces résultats ne permettent pas de discriminer entre des régions ouvertes de la chromatine sans nucléosome des régions où des nucléosomes sont apposés. Malheureusement, bien que les données de séquençage soient de bonnes qualité, ces résultats reflètent probablement des aléas techniques ne permettant pas d'obtenir un enrichissement des lectures au niveau des sites d'initiation de la transcription comme attendu ; des mises au point complémentaires devront être réalisées.

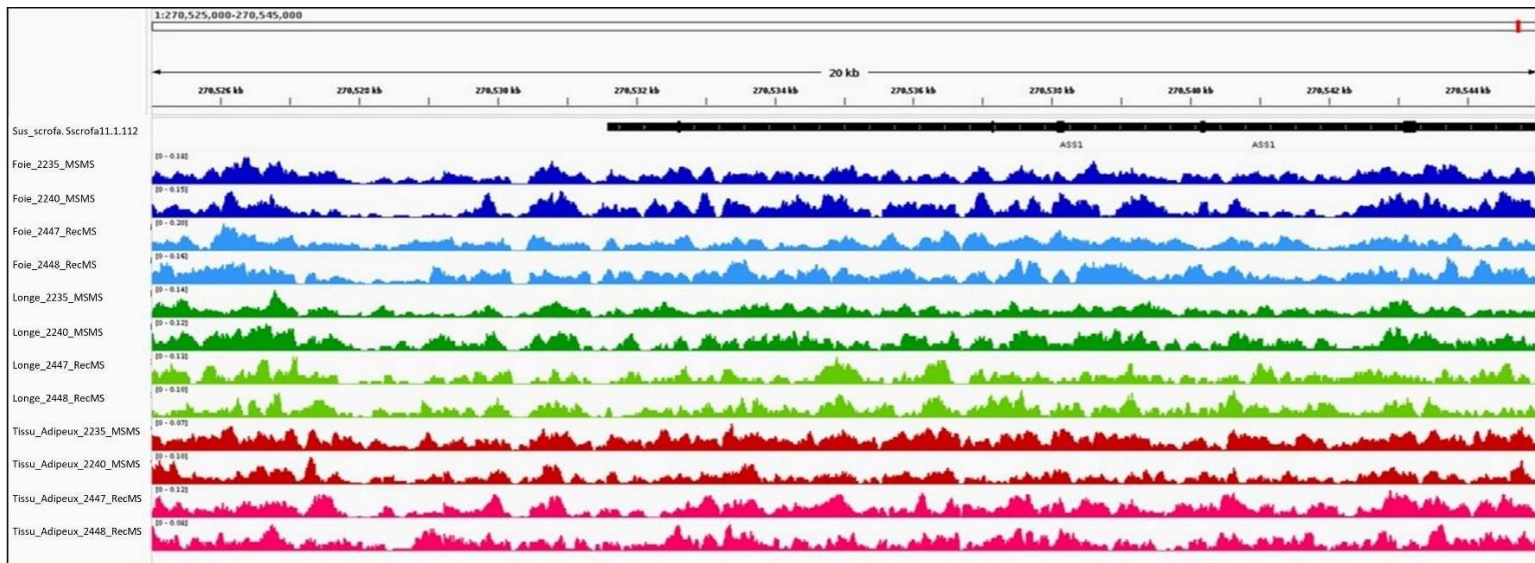


Figure 19 : Représentation IGV focalisée sur le gène *ASS1* pour chaque échantillon. Bleu : foie, vert : longe, rouge : tissu adipeux, foncé : génotype Meishan/Meishan, clair : Recombinant/Meishan.

Discussion/conclusion

Discussion

L'objectif de mon stage était de mener une analyse fonctionnelle d'une région localisée sur le bras long du chromosome 1 porcine et contenant un QTL de croissance et d'adiposité. Pour ce faire, j'ai travaillé sur des échantillons de trois tissus issus d'animaux homozygotes (MS/MS) ou hétérozygotes (Rec/MS) au QTL. Pour ce QTL, l'allèle d'origine Meishan est favorable et récessif ce qui signifie que les animaux homozygotes MS/MS présentent un taux de gras plus important. Tout d'abord, des analyses d'expression des gènes par RT-qPCR de la région d'intérêt ont été menées. Puis, des expériences d'accessibilité de la chromatine par ATAC-seq ont été réalisées.

L'analyse de l'expression génique par RT-qPCR est une approche à envisager pour identifier un gène potentiellement impliqué dans un phénotype donné. En effet, une mutation est susceptible d'altérer significativement le niveau d'expression du transcrit. Ainsi il est pertinent de quantifier l'expression de ces gènes par RT-qPCR pour détecter d'éventuelles régulations transcriptionnelles liées à la mutation. Nos résultats montrent que seuls les gènes *NCSI* (en dehors de l'intervalle minimum du QTL) et *HMCN2* (dans l'intervalle du QTL) semblent différenciellement exprimés entre des individus Rec/MS et des individus MS/MS, et seulement dans le foie. Toutefois, lors de la validation des données, ces 2 gènes présentaient des Ct très tardifs (34 Ct et 36 Ct respectivement), ce qui réduit la sensibilité de la méthode, rendant l'interprétation des résultats délicate. Pour valider les résultats, il serait préférable d'effectuer une nouvelle extraction d'ARN total. De plus, au vu de l'effet de l'allèle MS sur le phénotype d'adiposité, nous faisons l'hypothèse que les individus MS/MS exprimeraient le gène effecteur de manière plus importante que les individus Rec/MS. Pour la longe, une expression légèrement plus élevée du gène *HMCN2* est observée mais en faveur des individus

Rec/MS. Cette tendance est également observée dans le tissu adipeux pour le gène *ASS1*. En conclusion, ces premiers résultats nécessitent d'être reproduits pour confirmer ou infirmer le gène responsable de l'effet du QTL. Pour cette étude, une approche par RNA-seq pourrait constituer une alternative intéressante à la RT-qPCR, en offrant à la fois une quantification plus précise de l'ARN, la détection de différents isoformes d'un même gène, ainsi que l'identification potentielle de longs ARN non codants, souvent impliqués dans la régulation de l'expression génique (Berge et al. 2019).

Si aucun différentiel d'expression n'était retrouvé par RT-qPCR ou RNA-seq, il est probable que l'hypothèse d'un variant régulateur soit incorrecte; les analyses génétiques préalables ayant menées à cette hypothèse devraient être certainement approfondies. Il se pourrait donc que la mutation impliquée n'affecterait pas les transcrits mais pourrait avoir un impact au niveau protéique, comme cela a été montré pour *BMP15*. En effet, Otsuka *et al* ont montré que plusieurs mutations non synonymes ont été identifiées pour le gène *BMP15* impliqué dans la fonction ovarienne de plusieurs races ovines. La maturation et la sécrétion de BMP15 se trouve alors affectées, perturbant son interaction avec GDF9 pour la formation d'un hétérodimère (Otsuka et al. 2011). Une autre hypothèse possible serait un variant modifiant l'épissage d'un gène présent dans le QTL. A titre d'exemple, une mutation altérant l'épissage du gène *MLPH* chez le lapin a été montrée comme responsable de la dilution de la couleur du pelage. En effet, la stabilité et l'accumulation des transcrits n'étaient pas modifiées entre les animaux ou non de la mutation. C'est le type de transcrit qui différait avec un transcrit normalement épissé chez les individus sauvages et un transcrit avec une rétention d'intron chez les individus mutés (Demars et al. 2018).

Afin de poursuivre nos analyses fonctionnelles de la région d'intérêt et pour confirmer ou non notre hypothèse d'un variant régulateur, les régions régulatrices actives ont été analysées. Le principe des expériences d'ATAC-seq repose sur l'utilisation de la transposase

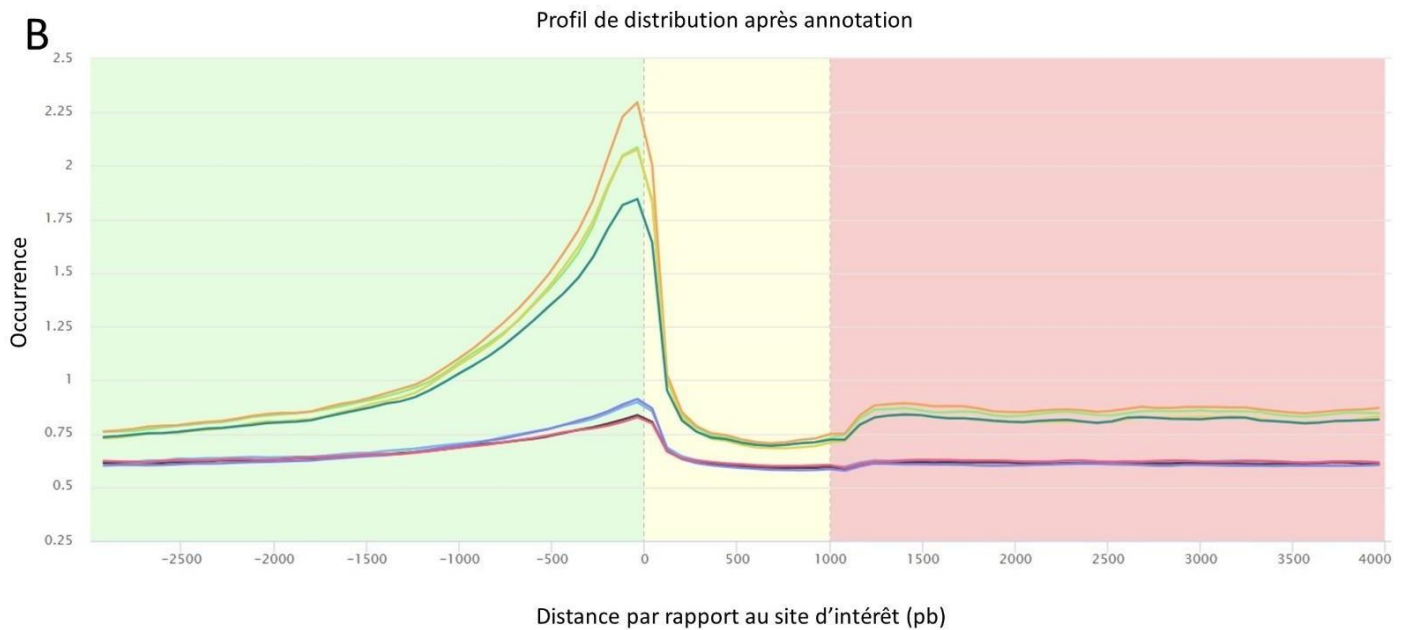
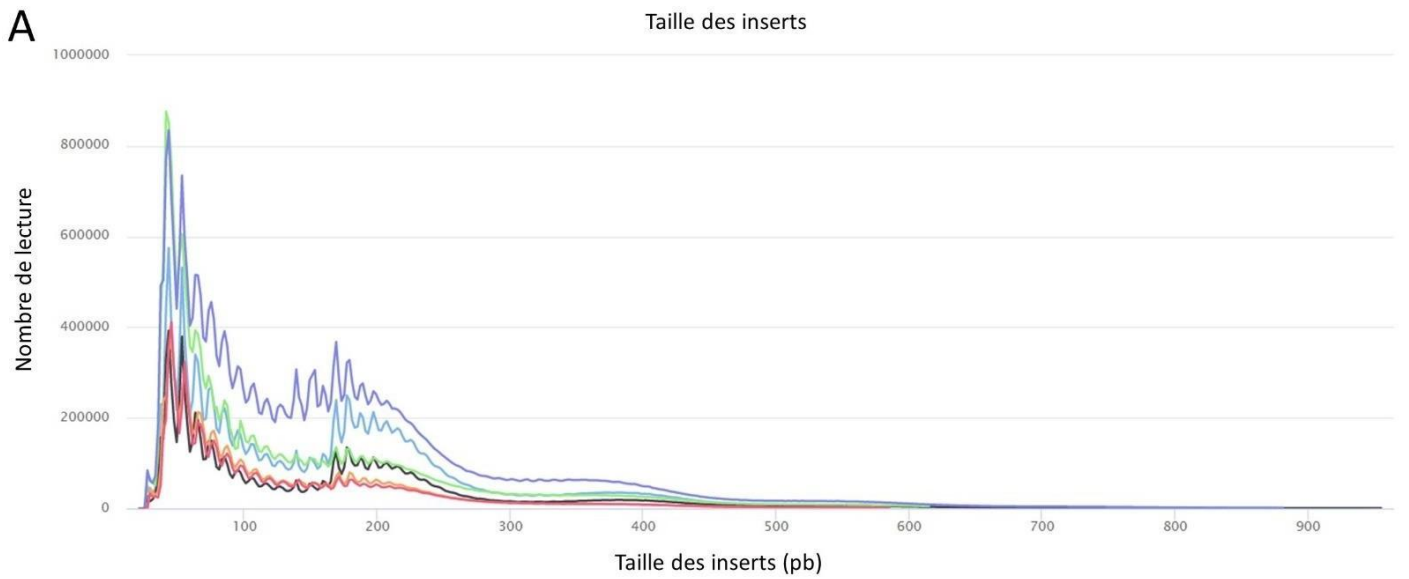


Figure 20 : Contrôles qualité attendu de l'analyse bio-informatique d'ATAC-seq. A. Distribution des tailles d'inserts attendue en paire de base : un pic à 50/60 pb et un pic autour de 150 pb (Picard Tools). **B.** Profil de distribution autour des TSS attendu. La zone verte correspond à la région promotrice, en amont du TSS. La zone jaune représente le corps du gène qui commence au TSS et fini au site de terminaison de la transcription (TES). La zone rouge correspond à la région en aval du gène (dDeptools)

Tn5 qui coupe l'ADN accessible et insère des adaptateurs de séquençage environ toutes les 60 pb. Dans les régions ouvertes de la chromatine sans nucléosome, la transposase Tn5 génère de nombreux fragments de 60 pb. Cependant, lorsqu'il y a la présence d'un nucléosome, la Tn5 effectue des coupures en amont et en aval, produisant des fragments de 146 pb, correspondant à la taille de l'ADN enroulé autour d'un nucléosome (**Figure 20A**). Dans les régions plus compactées de la chromatine, des fragments de tailles variables indiquent la présence de plusieurs nucléosomes successifs (290 à 800 pb). Nos résultats ont abouti à un protocole optimisé pour l'obtention de noyaux en quantité et qualité suffisantes pour des expériences d'ATAC-seq, comme montré par l'absence de lectures alignées sur le génome mitochondrial (**Figure 18A**). Cependant, les autres analyses post-séquençage ne montrent malheureusement pas un enrichissement des lectures au niveau du TSS pour l'ensemble du génome comme normalement attendu dans ces expériences (**Figure 20B**). Alors qu'un seul échantillon de tissu adipeux présente un profil caractéristique d'enrichissement au niveau des TSS, celui-ci ne semble pas suffisant par rapport au reste du génome (**Figure 18C**), suggérant un bruit de fond *via* des lectures alignées tout le long du génome. Ceci est confirmé d'une part par la comparaison de nos résultats préliminaires à des données obtenus dans le cadre du consortium FAANG, qui montre des tailles d'inserts courtes et des lectures réparties de manière homogène sur tout la région génomique (**Figure 21**). D'autre part, les profils des tailles des inserts mettent en évidence une variabilité entre les échantillons, avec 2 types de profils observés, le premier présentant un pic unique aux alentours de 50/60 pb et le second montrant un pic allant de 50 à 200 pb. L'absence des pics caractéristiques correspondant aux régions sans/avec nucléosomes, respectivement à 50/60 pb et 150 pb, combinée à la répartition des fragments sur l'ensemble du génome, semble indiquer une tagmentation excessive ou une dégradation de l'ADN.

Malheureusement, les résultats obtenus en ATAC-seq ne permettent pas d'interpréter biologiquement les résultats. Ces expériences illustrent la difficulté de mises au point pour des

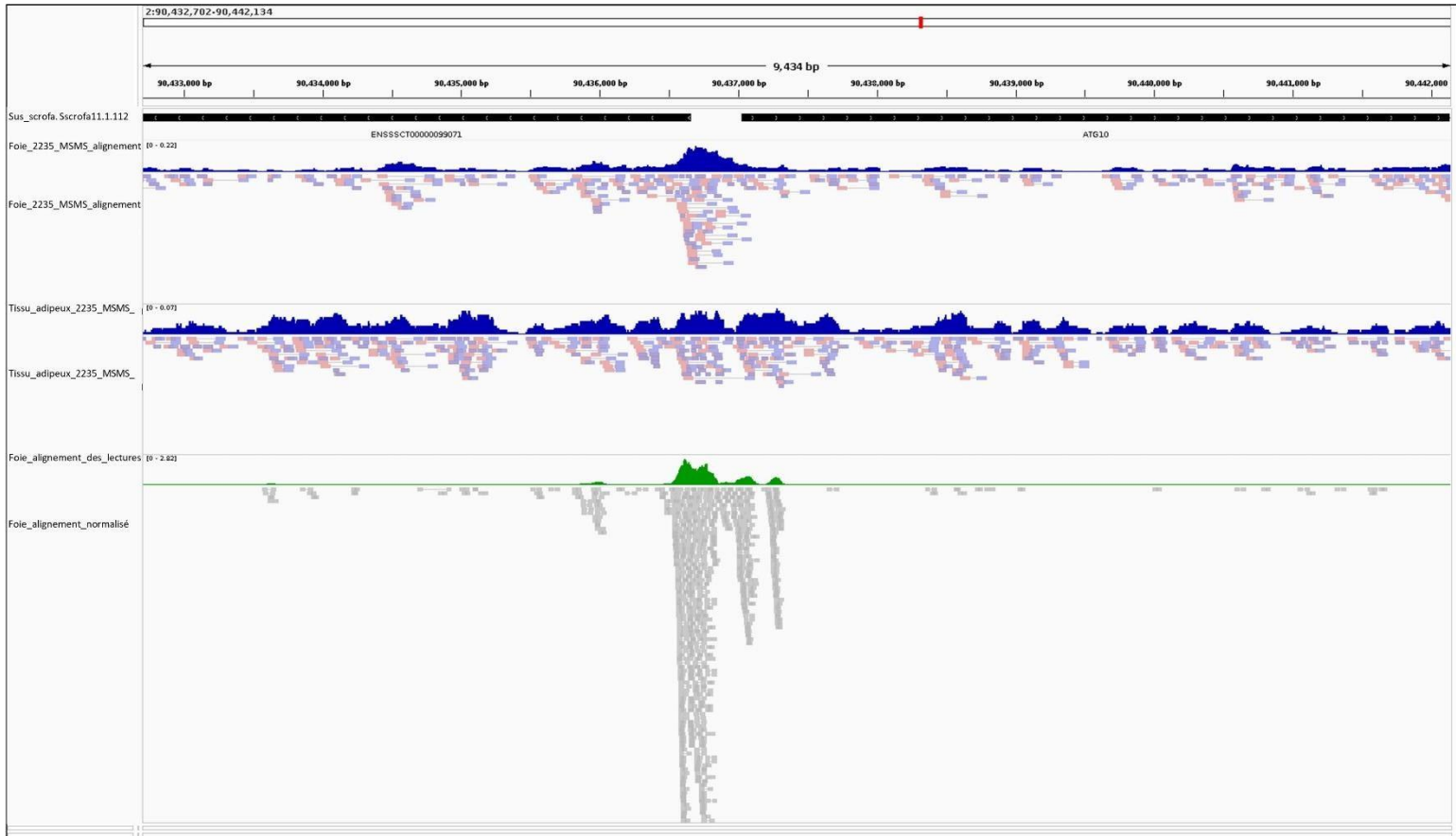


Figure 21 : Représentation IGV d'une région du chromosome 2. Lectures brutes et normalisées pour un échantillon de foie (piste 1), de tissu adipeux (piste 2) et de foie venant des données FAANG (piste 3).

technologies -omiques car seules les analyses post-séquençage permettent d'évaluer la qualité des données. Afin de tester l'hypothèse d'une sur-tagmentation, il serait pertinent de réaliser une cinétique de tagmentation. Les tailles des fragments obtenus à chaque temps pourraient être vérifiées. En cas de tagmentation optimale, un pic avec une intensité élevée est attendu aux alentours de 100 pb, correspondant aux régions ouvertes avec les adaptateurs, ainsi qu'un épaissement vers 180 pb, représentatif des régions associées à un nucléosome. Des pics à des tailles plus importantes, reflétant la présence de plusieurs nucléosomes, sont également attendus. Par ailleurs, pour vérifier l'hypothèse d'une dégradation d'échantillon, une extraction d'ADN, des tissus utilisés, et une migration sur gel pourraient être réalisées.

Conclusion – Perspectives

Pour conclure, les travaux que j'ai menés dans le cadre de mon stage visaient à appréhender d'un point de vue fonctionnel et de façon ciblée les effets d'un QTL. Un différentiel d'expression a été observé dans le foie pour quelques gènes lors de l'analyse de l'expression génique, bien que ces résultats restent à confirmer. Concernant l'ATAC-seq, les modifications apportées à l'étape d'isolement des noyaux ont été efficaces. Toutefois, la préparation des librairies nécessite quelques optimisations.

La réalisation d'expériences en laboratoire n'est jamais une tâche facile, cela demande à la fois un investissement important en temps pour l'optimisation et en ressources financières. Pour ces raisons, des alternatives utilisant des méthodes statistiques ou encore de l'intelligence artificiel (IA) sont de plus en plus développées et utilisées afin d'orienter l'identification de variants causaux (Kelley 2020). Ces outils permettent de limiter les expériences en ciblant les expériences les plus probables, telles que la modification d'histone ou la liaison de facteurs de transcription. Bien que l'IA facilite la démarche et réduise le nombre d'essais nécessaires en laboratoire, il demeure indispensable de valider expérimentalement les résultats obtenus par modélisation.

Bibliographie

Abiola O, Angel JM, Avner P, Bachmanov AA, Belknap JK, Bennett B, Blankenhorn EP, Blizard DA, Bolivar V, Brockmann GA et al. (2003) The nature and identification of quantitative trait loci: a community's view. *Nat Rev Genet* 4:911–916.

Aspenström P (2010) Formin-binding proteins: Modulators of formin-dependent actin polymerization. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803:174–182.

Berge KV den, Hembach KM, Soneson C, Tiberi S, Clement L, Love MI, Patro R and Robinson MD (2019) RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annual Review of Biomedical Data Science* 2:139–173.

Bidanel JP and Milan D (2000) La recherche de QTL à l'aide de marqueurs : résultats chez le porc. *INRAE Productions Animales* 13:223–228.

Bidanel J-P, Milan D, Iannuccelli N, Amigues Y, Boscher M-Y, Bourgeois F, Caritez J-C, Gruand J, Le Roy P, Lagant H et al. (2001) Detection of quantitative trait loci for growth and fatness in pigs. *Genetics Selection Evolution* 33:289.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY and Greenleaf WJ (2013) Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *Nat Methods* 10:1213–1218.

Davis-Smyth T, Duncan RC, Zheng T, Michelotti G and Levens D (1996) The Far Upstream Element-binding Proteins Comprise an Ancient Family of Single-strand DNA-binding Transactivators *. *Journal of Biological Chemistry* 271:31679–31687.

Demars J, Iannuccelli N, Utzeri VJ, Auvinet G, Riquet J, Fontanesi L and Allain D (2018) New Insights into the Melanophilin (MLPH) Gene Affecting Coat Color Dilution in Rabbits. *Genes* 9:430.

Demars J and Mourad R Cross-species use of neural networks trained on human and mouse sequences to predict chromatin regulation from pig sequences for genetics applications.

Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y and Segal E (2012) ENCODE explained. *Nature* 489:52–54.

Fève K Cartographie fine et caractérisation d'un QTL localisé sur le chromosome1 porcin influençant l'adiposité des animaux. EPHE. INRAE.

Fischer TT, Nguyen LD and Ehrlich BE (2021) Neuronal calcium sensor 1 (NCS1) dependent modulation of neuronal morphology and development. *The FASEB Journal* 35:e21873.

FranceAgriMer (2025) Viande porcine. Fiche filière.

Giuffra E, Tuggle CK and Consortium F (2019) Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annual Review of Animal Biosciences* 7:65–88.

Grandi FC, Modi H, Kampman L and Corces MR (2022) Chromatin accessibility profiling by ATAC-seq. *Nat Protoc* 17:1518–1552.

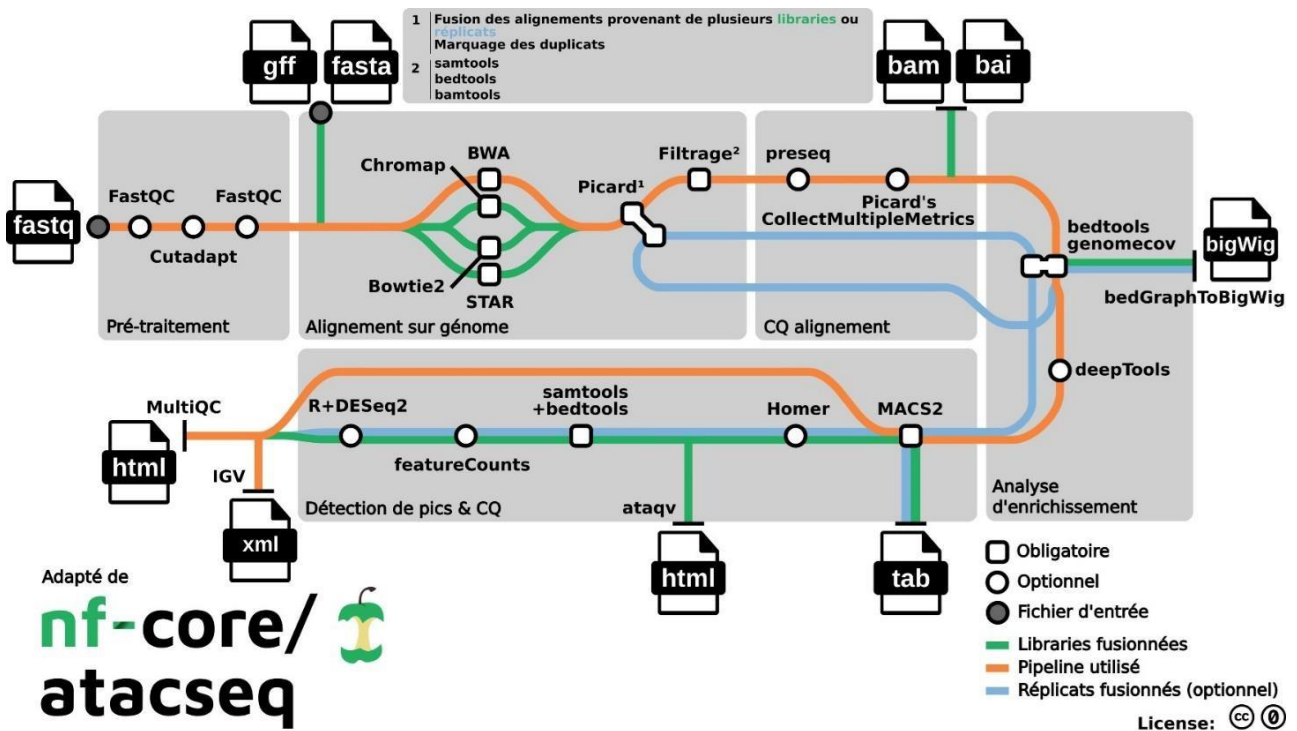
Kelley DR (2020) Cross-species regulatory sequence activity prediction. *PLoS Comput Biol* 16:e1008050.

- Krażek M, Wojciechowicz T, Fiedorowicz J, Strowski MZ, Nowak KW and Skrzypski M (2024) Neuronostatin regulates proliferation and differentiation of rat brown primary preadipocytes. *FEBS Letters* 598:1996–2010.
- Lee BM and Mahadevan LC (2009) Stability of histone modifications across mammalian genomes: Implications for ‘epigenetic’ marking. *Journal of Cellular Biochemistry* 108:22–34.
- Legault C, Caritez JC, Dupont C and Gogue J (1983) L’expérimentation sur le porc chinois en France I. Performances de reproduction en race pure et en croisement. *Génétique, sélection, évolution* 15:225.
- Li Z, Schulz MH, Look T, Begemann M, Zenke M and Costa IG (2019) Identification of transcription factor binding sites using ATAC-seq. *Genome Biology* 20:45.
- Lu Z, Ricci WA, Schmitz RJ and Zhang X (2018) Identification des éléments *cis* -régulateurs par la structure de la chromatine. *Current Opinion in Plant Biology* 42:90–94.
- Milan D, Bidanel J-P, Iannuccelli N, Riquet J, Amigues Y, Gruand J, Le Roy P, Renard C and Chevalet C (2002) Detection of quantitative trait loci for carcass composition traits in pigs. *Genetics Selection Evolution* 34:705.
- Mourot J, Kouba M and Salvatori G (1999) Facteurs de variation de la lipogenèse dans les adipocytes et les tissus adipeux chez le porc. *INRAE Productions Animales* 12:311–318.
- Otsuka F, McTavish KJ and Shimasaki S (2011) Integral role of GDF-9 and BMP-15 in ovarian function. *Molecular Reproduction and Development* 78:9–21.
- Quéméneur K, Marion A, Devine M and Le Gall M (2023) 29. Relationship between the body condition and lactating performance of sows. *Animal - science proceedings* 14:668.
- Rabinovich S, Adler L, Yizhak K, Sarver A, Silberman A, Agron S, Stettner N, Sun Q, Brandis A, Helbling D et al. (2015) Diversion of aspartate in ASS1-deficient tumours fosters de novo pyrimidine synthesis. *Nature* 527:379–383.
- Robert-Granié C, Legarra A and Ducrocq V (2011) Principes de base de la sélection génomique. *INRAE Productions Animales* 24:331–340.
- Schwob S, Lebreton B and Louveau I (2020) Adiposité et génétique chez le porc : état des lieux et nouveaux enjeux pour la qualité des produits: (Full text available in English). *INRAE Productions Animales*. doi: 10.20870/productions-animales.2020.33.1.3112
- Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF et al. (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507:371–375.
- Taylor G, Roesse G and Hermes S Breeds of pigs - Large White.
- Umlauf D and Mourad R (2019) The 3D genome: From fundamental principles to disease and cancer. *Seminars in Cell & Developmental Biology* 90:128–137.
- Wang Z, Wei G-H, Liu D-P and Liang C-C (2007) Unravelling the world of cis-regulatory elements. *Med Bio Eng Comput* 45:709–718.
- Welcker D, Stein C, Feitosa NM, Armistead J, Zhang J-L, Lütke S, Kleinridders A, Brüning JC, Eming SA, Sengle G et al. (2021) Hemicentin-1 is an essential extracellular matrix component of the dermal–epidermal and myotendinous junctions. *Sci Rep* 11:17926.

Zarnovican MH (2024) Production et commerce mondiaux de viande porcine en 2023. BioCLips 32:no3.

Zeyland J, Woźniak A, Gawrońska B, Juzwa W, Jura J, Nowak A, Słomski R, Smorąg Z, Szalata M, Mazurek U et al. (2014) Double transgenic pigs with combined expression of human α 1,2-fucosyltransferase and α -galactosidase designed to avoid hyperacute xenograft rejection. Arch Immunol Ther Exp (Warsz) 62:411–422.

Annexes



Annexe 1 : Schéma du pipeline **nf-core-atacseq**. Le chemin orange indique les différentes étapes du pipeline utilisées (modifié d'après **nf-core/atacseq** (<https://nf-co.re/atacseq/2.1.2/>)).