

## SUJET DE STAGE DE MASTER 2 2020-2021

### RECOMBINAISON MEIOTIQUE ET DIVERSITE DU CONTENU EN ELEMENTS TRANSPOSABLES DANS LE GENOME DE LA SOURIS.

#### Lieu

Institut de Génétique Humaine de Montpellier, UMR 9002, 141 rue de la Cardonille, 34396 Montpellier cedex 5  
Equipe Méiose et recombinaison (direction Bernard de Massy)

<https://www.igh.cnrs.fr/fr/recherche/departements/dynamique-du-genome/5-meiose-et-recombinaison>

#### Encadrant

Julie Clément

✉ [julie.clement@igh.cnrs.fr](mailto:julie.clement@igh.cnrs.fr)

☎ +33 4-34-35-99-72

#### Contexte du projet et objectifs

Une des caractéristiques des génomes est la présence d'éléments transposables (ET). Ces éléments, aussi appelés éléments génétiques mobiles, peuvent se répliquer, se déplacer dans le génome en s'excisant puis se re-intégrant. Ils sont autonomes, ce sont des éléments dit égoïstes, des sortes de parasites de génome et dont la fonction pour les cellules est encore mal comprise. Leur proportion dans les génomes est variable entre les espèces (de 2 à 80%) et chez la souris, environ 50% du génome est constitué d'ET. Certains ont été fixés dans les génomes au cours de l'évolution, et d'autres sont encore actifs et mobiles. La diversité du contenu en ET est présente au sein même des espèces et une étude chez 18 sous-espèces de souris<sup>1</sup> a permis d'identifier plus de 100 000 variants d'insertion. Ce grand nombre de variants pose des questions majeures sur la stabilité du génome comme par exemple lors de la méiose.

Chez les organismes se reproduisant par voie sexuée, la recombinaison méiotique est un mécanisme essentiel pour la fertilité et un moteur de diversité génétique. Lors de la première division méiotique, des centaines de cassures de l'ADN double-brin se produisent puis se réparent en utilisant une séquence d'ADN du chromosome homologue comme modèle. Ces cassures ont lieu dans des régions bien spécifiques du génome. Chez la souris, comme chez de nombreux Mammifères, elles sont déterminées par la fixation de la protéine PRDM9<sup>2,3</sup> qui reconnaît une séquence d'ADN spécifique. La diversité des allèles de *Prdm9* de la souris est remarquable<sup>4</sup>. En conséquence, la cartographie des cassures méiotiques est totalement différente entre différents génotypes de souris<sup>5</sup>.

La distribution de ces cassures par rapport aux différents ET n'a été que très peu étudiée, surtout pour des raisons méthodologiques. De nombreuses questions évolutives et moléculaires sont donc à explorer. En particulier, on ne connaît pas la contribution des ET à l'activité de cassures chez différents génotypes de souris ni même si toutes les copies d'un élément ont la même activité de recombinaison. Existe-t-il des ET systématiquement ciblés par différents PRDM9 ou au contraire évités ? Aussi, la présence de cassures dans ces éléments répétés pose la question du choix de la copie pour la réparation : si une copie en position non-homologue est utilisée (c'est-à-dire une copie distante, située à un autre locus), il y a un risque mutagène voire un risque de réarrangement chromosomique. Ceci justifie d'autant plus de caractériser l'étape précoce de la localisation des cassures dans ces éléments du génome.

Ce projet de recherche vise à étendre nos connaissances sur l'activité de recombinaison méiotique dans les éléments transposables chez la souris. En particulier, lors de ce stage, nous aborderons les questions suivantes :

1/ Quelle est la diversité des ET ciblés par PRDM9 ? Pour cela, nous allons annoter les sites de cassures chez plusieurs sous-espèces de souris et décrire les différents ET (familles, classes, âge des éléments) ciblés par divers allèles de *Prdm9*.

2/ Quelle est la conséquence du polymorphisme des ET sur l'activité de recombinaison ? Pour cela, nous allons analyser les sites de cassures chez des souris hybrides, en particulier au niveau des sites polymorphiques entre les deux génomes parentaux.

#### Méthodes

La position des sites de cassures est permise grâce à un protocole modifié de ChIP-seq<sup>6,7</sup>. Dans le cadre du travail proposé ici, l'ensemble des expérimentations a déjà été réalisé, et les données de NGS (Next Generation

Sequencing) sont déjà disponibles et publiées. Le travail du (de la) candidat(e) consistera en l'analyse spécifique de ces données de ChIP-seq au niveau des ET.

La nature même des ET constitue un réel challenge pour l'analyse de données NGS<sup>8,9</sup>. D'un côté, à cause de leur caractère répété, l'alignement des séquences provenant des ET sur le génome de référence est un réel défi méthodologique qui requiert l'utilisation ou le développement d'outils dédiés<sup>10</sup>. D'un autre côté, les génomes de référence sont imparfaitement assemblés dans ces régions répétées et les annotations des ET incomplètes, d'autant plus que de nouvelles insertions peuvent apparaître. Il est donc impossible d'avoir un répertoire exhaustif du contenu et de la localisation des ET pour le génome de la souris de référence ou des autres génotypes.

Durant ce stage, nous travaillerons alors sur ces deux aspects. Tout d'abord, la construction d'un catalogue des ET sera réalisée et inclura les polymorphismes de chaque génotype étudié. Ensuite, pour pallier aux défauts des méthodes classiques d'alignement, nous utiliserons alors au moins deux approches : la réassignation des multi-mappers et l'alignement local sur une collection d'ET. Ce travail nécessitera le développement d'un pipeline d'analyse (sous snakemake ou nextflow) et de scripts dédiés, permettant entre autre la manipulation des données, leur synthèse et leur visualisation.

Sur le plan technique, ce projet permettra une formation approfondie en analyse de données haut débit, en bioinformatique, en programmation et manipulation avancée de tableaux. Les principes de la science reproductible et du travail collaboratif seront également appliqués.

### Résultats attendus

- Annotation et comparaison des ET aux points chauds de diverses sous-espèces de souris
- Evaluation du polymorphisme nucléotidique et d'insertion/délétion des ET entre les différentes sous-espèces
- Analyse de l'impact de la présence d'ET sur l'activité des points chauds
- Analyse de l'impact du polymorphisme sur l'activité des points chauds
- Synthèse sur la contribution des ET à la diversité des cartes de recombinaison chez la souris

### Profil et compétences recherchées

- Solide formation en bioinformatique et en analyse de données génomiques
- Maîtrise de l'environnement UNIX
- Programmation en bash et Python ou R.
- Intérêt marqué pour l'évolution des génomes, la diversité du monde vivant et les processus moléculaires à l'origine de cette diversité.
- Rigueur et organisation dans le travail indispensable
- Curiosité et dynamisme

1. Nellåker, C. *et al.* The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* **13**, R45 (2012).
2. Baudat, F. *et al.* PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science (80-. ).* **327**, 836–840 (2010).
3. Grey, C. *et al.* In vivo binding of PRDM9 reveals interactions with non-canonical genomic sites. *Genome Res.* 1–12 (2017). doi:10.1101/gr.217240.116.Freely
4. Buard, J. *et al.* Diversity of Prdm9 Zinc Finger Array in Wild Mice Unravels New Facets of the Evolutionary Turnover of this Coding Minisatellite. *PLoS One* **9**, e85021 (2014).
5. Smagulova, F., Brick, K., Pu, Y., Camerini-otero, R. D. & Petukhova, G. V. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev.* 266–280 (2016). doi:10.1101/gad.270009.115.4
6. Khil, P. P., Smagulova, F., Brick, K. M., Camerini-Otero, R. D. & Petukhova, G. V. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res.* **22**, 957–965 (2012).
7. Brick, K., Pratto, F., Sun, C. Y., Camerini-Otero, R. D. & Petukhova, G. *Analysis of Meiotic Double-Strand Break Initiation in Mammals. Methods in Enzymology* **601**, (Elsevier Inc., 2018).
8. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2012).
9. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
10. Teissandier, A., Servant, N., Barillot, E. & Bourc'His, D. Tools and best practices for retrotransposon analysis using high-throughput sequencing data. *Mob. DNA* **10**, 1–12 (2019).