

## **Stage M1: Détection de variants et génération de séquence de référence pour des lignées cellulaires (murin, humain et rat)**

**Niveau :** M1 bioinformatique      **Durée:** 4 mois

**Date :** selon la formation (Avril/Mai?)

**Encadrement technique:** Céline Noirot (MIAT / INRAE Toulouse Castanet)

**Encadrement scientifique :** Jacques Dupuy (Toxalim / INRAE Toulouse Saint-martin)

**Localisation :** Inrae Castanet (mais sera probablement en TT)

### **Profil recherché :**

Connaissances en génomique et outils pour le traitement de données génomiques (Samtools, BWA, GATK, VCFtools)

Maîtrise de **Unix** et d'un ordonnanceur (eg: SLURM/SGE)

Maîtrise de l'anglais technique

### **Contexte scientifique:**

Le projet consiste à traiter des données de séquençage illumina de lignées cellulaires pour identifier des mutations spécifiques aux lignées.

Les lignées cellulaires épithéliales coliques murines et de fibroblastes murins vont permettre de confirmer la mutation d'un gène clé dans le développement des cancers colorectaux (Adenomatous Polyposis Coli = APC) pour les lignées Apc min/+ par rapport aux lignées témoin (Apc +/+). De plus, dans 2 clones issus des lignées cellulaires épithéliales et invalidées par la technique Crispr/Cas9 pour un facteur de transcription majeur dans le système de protection contre le stress oxydatif (Nuclear factor (erythroid-derived 2)-like 2 = Nrf2), nous identifierons les modifications au niveau du gène nrf2 (lignée Apc +/+ Nrf2 -/- et apc min/+ nrf 2-/-). Dans un second temps, nous déterminerons les mutations du gène APC dans une lignée épithéliale colique humaine normale (NCM 356) invalidée pour ce gène via la technique Crispr/Cas9 en comparaison avec la lignée normale. Enfin, nous déterminerons les mutations dans des échantillons (foies) issus de rats invalidés pour le gène Nrf2 en comparaison avec des animaux normaux.

Au final, cette approche permettra de créer des bibliothèques de référence pour les lignées murines et humaines ainsi que d'identifier l'ensemble des modifications géniques associées à la délétion du gène nrf2 chez des rats.

### **Traitement à réaliser:**

La détection de variants a déjà été réalisée grâce au pipeline Sarek en utilisant les génomes de références publics. Le stage consistera donc à

- filtrer des variants selon les conditions
- générer d'une nouvelle référence murine
- détecter à nouveau les variants avec le pipeline Sarek en utilisant la nouvelle référence. Évaluer la qualité de ces nouveaux variants.
- exploration des données répliquées pour comprendre d'où cela provient exactement (mutation des cellules?)

**Contact :** [celine.noirot@inrae.fr](mailto:celine.noirot@inrae.fr) & [jacques.dupuy@inrae.fr](mailto:jacques.dupuy@inrae.fr)