

# Identification de relations enhancer/gène dans les génomes

## Proposition de stage niveau M2 (printemps 2021)

Unités IRSD, INSERM et MIAT, INRAE, Toulouse

Pour postuler envoyer CV, lettre de motivation et dernier relevé de notes à

[sarah.djebali@inserm.fr](mailto:sarah.djebali@inserm.fr) et [nathalie.vialaneix@inrae.fr](mailto:nathalie.vialaneix@inrae.fr)

### Problématique du stage

Pour mieux comprendre les maladies génétiques complexes, il est important de mieux comprendre le fonctionnement du génome, et en particulier les éléments et molécules impliqués dans la régulation de l'expression des gènes.

Aujourd'hui de très nombreuses données fonctionnelles sont disponibles pour investiguer cette question, telles que celles issues de grands consortia internationaux comme ENCODE, REMC ou FANTOM.

Ce stage vise à utiliser ces données pour mieux comprendre les relations régulatrices de type enhancer/gène afin de mieux pouvoir les identifier automatiquement dans les génomes.

### Travail à effectuer

Après avoir dressé un état de l'art des méthodes existantes, qui utilisent le plus souvent le *machine learning* (random forest, réseaux de neurones), il s'agira dans un premier temps de constituer des ensembles de relations enhancer/gène de référence pour l'évaluation non biaisée des méthodes. Il s'agira ensuite de normaliser les méthodes existantes afin de les rendre comparables entre elles (utilisation des mêmes données de validation, des mêmes règles heuristiques internes, ...), avant de les évaluer sur les données de référence.

### Conditions du stage

- durée: 4 à 6 mois
- localisation :  
[IRSD](#) (Institut de Recherche en Santé Digestive) - INSERM U1220  
Equipe génétique et régulation du métabolisme du fer
- rémunération : taux légal (environ 550 euros par mois)
- encadrement : [Sarah Djebali](#) et [Nathalie Vialaneix](#)
- contact : [sarah.djebali@inserm.fr](mailto:sarah.djebali@inserm.fr) et [nathalie.vialaneix@inrae.fr](mailto:nathalie.vialaneix@inrae.fr)

### Profil recherché

- Master 2 en bioinformatique ou biostatistiques ou équivalent en école d'ingénieur
- bien connaître UNIX et la ligne de commande
- savoir programmer
- notions en statistique (et très bonne maîtrise de R ; une maîtrise de python serait un plus) et en machine learning
- avoir déjà manipulé des données issues du séquençage haut-débit et/ou avoir des connaissances en biologie moléculaire seraient un plus