

## FORMULAIRE STAGE Recherche-M2 BBSG

**Titre du stage:** Analyses bioinformatiques de données de séquençage pour l'étude de la replication des enhancers et des promoteurs au cours du cycle cellulaire.

**Laboratoire :** TAGC/INSERM U1090,  
163, avenue de Luminy, case 928  
13288 Marseille cedex 09

**Equipe :** Bioinformatique des régions régulatrices  
**Maître de stage :** Benoit Ballester  
**E-mail :** [benoit.ballester@inserm.fr](mailto:benoit.ballester@inserm.fr)  
**Téléphone :** 04 91 82 87 28  
**Site :** <http://goo.gl/1V82gA>

Je propose un stage M2R en bioinformatique pour tout étudiant du BBSG qui souhaite:

- apprendre à manipuler et analyser des données NGS (Chip-seq, RNA-seq....)
- analyser les régions non-codantes comme les origines de réplifications, les enhancers
- programmer des outils et/ou des pipelines d'analyse en Python, Perl, R, autres langages.
- développer des méthodes originales d'analyse des données NGS
- rechercher des motifs de fixations dans ces régions
- travailler sur les serveurs de calculs du groupe bioinformatique du TAGC.

J'invite les étudiants intéressés par ce stage en bioinformatique à prendre contact au plus vite.

### Introduction:

Mes travaux sont axés sur l'étude des régions non-codantes des génomes (origines de réplifications) et des régions régulatrices (*enhancers*) qui fixent des facteurs de transcription. Les techniques de séquençage à haut débit (NGS) ont permis depuis quelques années de découvrir ces régions régulatrices, et l'importance des régions non-codantes autrefois appelés « *junk dna* ». Ces régions sont nombreuses dans le génome et l'objectif global du stage ainsi que la thèse est d'élucider spécifiquement les enhancers et des promoteurs dans le contexte de la répllication de l'ADN grâce à des approches bioinformatiques.

### Descriptif du stage:

D'une manière générale le stage consistera à utiliser des approches bioinformatiques pour d'une part intégrer des données issues du séquençage à haut débit (ChIP-seq, RNA-seq, DNase-seq) afin de déterminer les spécificités de la répllication des promoteurs et des enhancers.

**Le projet pourra s'adapter en fonction des centres d'intérêts et des compétences particulières du candidat (ex: orientation programmation et/ou analyse des données) ainsi que de l'avancement des projets au moment du stage.**

Ce sujet fait parti d'une thématique innovante et compétitive dans le contexte scientifique international. Ce sujet peut mener à un ou des sujets de thèses en bioinformatique et génomique.

### Environnement du stage:

Le ou la stagiaire sera accueilli au sein du laboratoire Inserm 1090 TAGC sur le campus de Luminy. L'étudiant(e) travaillera entouré de chercheurs en bioinformatique, ingénieurs bioinformaticiens et étudiants en thèse BBSG. Le TAGC est un laboratoire 50% bioinformatique, 50% génomique. Il permettra au stagiaire d'intégrer une équipe de recherche pluridisciplinaire, de travailler sur des

clusters de calculs et d'améliorer ses compétences générales en informatique et génomique. Pour plus d'informations, me contacter directement.

### Encadrement des étudiants BBSG/Autres :

Période	Type	Sujet
2016	Stage : M2R BBSG	Analyses bioinformatiques des régions régulatrices : Transcription des enhancers
2016	Stage : M1 BBSG	Catalogue des régions régulatrices chez <i>Drosophila melanogaster</i>
2016	Stage : PIB BBSG	Développement d'un programme pour calculer des cartes de densités issues de données de séquençage a haut débit
2011-2015	Thèse BBSG Bioinformatique	Etude de la complexité des éléments cis-régulateurs chez les mammifères en utilisant des approches a haut débit
2014	Stage : SupBioTech Paris	Développement de méthodes pour prioriser les régions régulatrices et les variant régulateurs dans le cancer du sein
2014	Stage : M2Pro BBSG	Développement du site <a href="http://tagc.univ-mrs.fr/remap/">http://tagc.univ-mrs.fr/remap/</a> Création d'un outil de calcul d'enrichissement
2013	Stage : PIB BBSG	Développement d'un package R pour le clustering des données de réplifications
2013	Stage : M1 BBSG	Analyse de données Chip-seq, création de pipeline

### Contact

Benoît Ballester, [benoit.ballester@inserm.fr](mailto:benoit.ballester@inserm.fr) (<http://goo.gl/1V82gA>)  
INSERM U1090, 163, Avenue de Luminy, Bâtiment INSERM, bloc 5,  
13288 MARSEILLE cedex 09. France

Christelle Cayrou, [christelle.CAYROU@unice.fr](mailto:christelle.CAYROU@unice.fr)  
INSTITUT DE RECHERCHE SUR LE CANCER ET LE VIEILLISSEMENT  
CNRS UMR 7284 - INSERM U 1081 - UNS  
28 avenue de Valombrose  
06107 Nice Cedex 02 – France

### Références:

Christelle Cayrou, **Benoît Ballester**, Isabelle Peiffer, Romain Fenouil, Philippe Coulombe, Jean-Christophe Andrau, Jacques van Helden, Marcel Méchali, The chromatin environment shapes DNA replication origin organization and defines origin classes, *Genome research* 2015

A Griffon, Q Barbier, J Dalino, J van Helden, S Spicuglia, **B Ballester**, Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape, *Nucleic acids research*, 2015

**B Ballester**, A Medina-Rivera, D Schmidt, M González-Porta, M Carlucci, et al. Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways, *eLife* 2014

Schmidt D\*, Wilson MD\*, **Ballester B\***, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, Talianidis I, Flicek P, Odom DT. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science*. 2010 Apr 8 (\* Co-first)

Schmidt D\*, Schwalie PC\*, Wilson MD, **Ballester B**, Gonçalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*. 2012 Jan 20;148(1-2):335-48.