

**FORMULAIRE STAGE Recherche-M2 BBSG**  
**(période de stage : du 5 janvier 2016 au 3 juillet 2016)**

**Titre du stage : Analyses bioinformatiques de données NGS pour l'étude des régions non-codantes et régulatrices du génome humain**

**Laboratoire (intitulé, adresse, site web) :** TAGC/INSERM U1090,  
163, avenue de Luminy, case 928  
13288 Marseille cedex 09

<b>Equipe :</b>	Bioinformatique des régions régulatrices
<b>Maître de stage :</b>	Benoit Ballester
<b>E-mail :</b>	<a href="mailto:benoit.ballester@inserm.fr">benoit.ballester@inserm.fr</a>
<b>Téléphone :</b>	04 91 82 87 28
<b>Site :</b>	<a href="http://goo.gl/1V82gA">http://goo.gl/1V82gA</a>

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, Chiapet, Hi-C, etc..)
- analyser les régions régulatrices (*enhancers*) et non-codantes dans le génome humain
- 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.

### Introduction:

Mes travaux sont axés sur l'étude des régions non-codantes des génomes 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 l'activité et la spécificité de ces régions régulatrices 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, Hi-C, Chia-PET, RNA-seq, DNase-seq) afin de déterminer l'activité des régions non-codantes du génome humain. Récemment il a été montré que les *enhancers* étaient transcrits (eRNA : *enhancer RNA*). Voici la/les questions que le candidat s'efforcera de répondre pendant son stage et thèse :

- Quelle est la fonction et les caractéristiques de cette transcription des régions régulatrices ?
- Est-il possible de prédire la transcription des *enhancers* grâce à des modèles bioinformatiques ?
- Comment les régions régulatrices interagissent avec les promoteurs des gènes qu'elles activent ?

Nos travaux récents (thèse BBSG) ont permis de révéler la complexité des régions régulatrices dans le génome (<http://goo.gl/UFQDd1>, <http://tagc.univ-mrs.fr/remap/>). En se basant sur ces travaux, nous sommes désireux aujourd'hui de décrire/déduire l'activité de ces régions régulatrices grâce à des approches bioinformatiques.

**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 a 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
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,  
13288 MARSEILLE cedex 09. France  
Bâtiment INSERM, bloc 5, Campus de Luminy  
Tel: 04 91 82 87 28

### Références:

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.