

# Stage de recherche en laboratoire

**Titre:** Epissage alternatif et progression tumorale

**Structure d'accueil :** Equipe « Epissage Alternatif et progression tumorale »,  
LBMC <http://www.ens-lyon.fr/LBMC>  
ENS de Lyon  
Adresse : 46, allée d'Italie, 69364 LYON CEDEX 07, FRANCE

**Tuteur scientifique:** Auboeuf, Didier, DR1 Inserm  
Téléphone: 04 72 72 87 94  
Courriel: didier.auboeuf@inserm.fr

**Mots clefs :** cancer ; épissage ; épigénétique ; RNASeq, Chip-Seq, HIC, base de données, pipelines et workflows d'analyse de données à large échelle.

**Description du projet :** L'épissage alternatif est une étape essentielle de l'expression des gènes car ce mécanisme augmente la diversité du protéome codé par un nombre limité de gènes. Ayant développé des compétences dans l'analyse à large échelle de l'épissage alternatif à partir de données RNAseq, nous avons montré que ce mécanisme contribue à la progression tumorale. Les cellules cancéreuses produisent des variants d'épissage et des isoformes protéiques qui contribuent à leur plasticité phénotypique.

Le projet proposé a pour objectif de mieux comprendre les mécanismes par lesquels les cellules cancéreuses produisent certains variants d'épissage. Nous pensons qu'il existe un lien entre l'organisation de la chromatine, l'organisation 3D du génome et l'épissage alternatif. Ainsi, une désorganisation de la topologie de l'ADN dans l'espace au cours de la progression tumorale pourrait perturber non seulement le niveau d'expression des gènes mais également la nature de leurs produits.

Pour adresser cette question, le candidat apprendra à utiliser des pipelines d'analyse de RNAseq (écrit en perl et R), de ChIPseq (en python, perl et R) mais aussi de HiC. Il devra également développer une base de données et son interface web (html5, javascript, php ou cgi-perl et MySQL). Cette base de données reprendra les résultats produits par les pipelines mentionnés ci-dessus. A partir de cette base de données, le candidat devra effectuer des analyses multivariées de différents types de données pour identifier des liens significatifs entre cooccurrence spatiale et corégulation de l'épissage. Le candidat développera également des outils de visualisation et de résumé statistique pour les large datasets produits par les pipelines.

Ainsi au cours de son stage le candidat apprendra à maîtriser des pipelines d'analyse bio-informatique en utilisant le gestionnaire de pipeline Nextflow. Il apprendra à maîtriser R et les modèles d'analyses multivariées. Il apprendra également à maîtriser des technologies de développement web et de base de données.

**Une bonne maîtrise de l'environnement UNIX est nécessaire, car le candidat aura à utiliser la grille de calcul de l'ENS pour ses analyses. Le candidat devra aussi utiliser git pour assurer le suivi de son travail.**

**Les candidats doivent être motivés par les questions biologiques et flexibles afin d'interagir avec les différents membres de l'équipe (biologistes et bioinformaticiens).**

## 5 Publications récentes de l'équipe:

- The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. Nucleic Acids Res. 2018

- Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data. Sci Rep. 2018

- Identification of protein features encoded by alternative exons using Exon Ontology. Genome Res. 2017

- The multiple functions of RNA helicases as drivers and switchers of the genetic information flow. Nature Rev. Mol. Cell Biol. 2017

- Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. Genome Res. 24: 511-21 (2014).