

## PROPOSITION DE STAGE

### M2 INFORMATIQUE ou BIONFORMATIQUE



**Titre :** D veloppement d'une nouvelle approche d'apprentissage automatique combinant le supervis  et le non-supervis  pour la classification des ARN non-codants.

**Mots-cl fs :** «Machine learning, R seaux de neurones, Cartes SOM, Apprentissage supervis , Apprentissage non-supervis , d veloppement logiciel, bioinformatique, pr diction d'ARN non-codants

#### Ma tres de stage :

*Fariza Tahi*, Professeur des Universit s, Equipe AROBAS, laboratoire IBISC, Universit  Paris-Saclay, Universit  d'Evry, Genopole

*Farida Zehraoui*, Maitre de Conf rences, Equipe AROBAS, laboratoire IBISC, Universit  Paris-Saclay, Universit  d'Evry

**Lieu du stage :** Laboratoire IBISC, IBGBI, Universit  d'Evry, 23 Boulevard de France, 91000 Evry.

**Dur e de stage :** 6 mois

**Contact :** fariza.tahi@univ-evry.fr

#### R sum  :

Les ARN non-codants (ARNnc) sont depuis quelques ann es au c ur de recherches particuli rement attrayantes et intensives, notamment dans le domaine biom dical. En effet, les ARNnc sont impliqu s dans un grand nombre de maladies, telles que le cancer, le diab te, les maladies neuro-d g n ratives, et suscitent un int r t croissant dans leur utilisation comme biomarqueurs ou comme cibles th rapeutiques. Gr ce aux nouvelles g n rations de s quenc age (NGS), le profilage d taill  des mol cules d'ARNnc a  t  consid rablement am lior e. Un grand nombre de classes et de types d'ARNnc ont  t  d couverts, mais beaucoup d'autres classes sont encore   d couvrir.

L'identification de ces ARNncs est donc un enjeu important, et avec les NGS qui g n rent des volumes consid rables de donn es RNAseq notamment, leur pr diction par des m thodes *in silico* est n cessaire. Ces m thodes permettront en effet d'identifier des ARNncs potentiels pouvant ensuite  tre valid s par des m thodes exp rimentales.

Le but de ce stage est de d velopper une m thode d'apprentissage automatique multi-sources combinant l'apprentissage supervis  et le non-supervis , afin d'identifier les ARNncs connus et de d couvrir de nouvelles classes, en utilisant des sources de donn es h t rog nes repr sentant diff rentes caract ristiques potentielles associ es   ces ARN : leur structure secondaire, la pr sence de certains motifs dans leur s quence, des caract ristiques  pig n tiques, etc. L'algorithme se basera sur des r seaux de neurones supervis s de type perceptron multi-couches (ou MLP) et des r seaux de neurones non-supervis s de type cartes auto-organisatrices (ou SOM pour Self Organizing Maps) pour permettre une meilleure visualisation et interpr tation des r sultats.

Ce stage fait suite aux travaux men s dans le cadre de la th se de Ludovic Platon (voir les publications ci-dessous), et s'inscrit dans la continuit  des travaux de pr diction des ARNncs et de leurs structures men s dans l' quipe AROBAS depuis plusieurs ann es. Ces travaux ont donn  lieu   plusieurs algorithmes et logiciels publi s dans des revues internationales et mis   disposition de la communaut  scientifique via le serveur web : <http://EvryRNA.ibisc.univ-evry.fr>

#### Bibliographie :

- L. Platon, F. Zehraoui, F.Tahi. Localized Multiple Sources Self-Organizing Map. International Conference on Neural Information Processing, ICONIP (3) 2018 : 648-659.
- L. Platon, F. Zehraoui, A. Bendahmane, F. Tahi. IRSOM, a reliable identifier of ncRNAs based on supervised Self-Organizing Maps with rejection. Bioinformatics, 34(17), i620-i628, 2018.
- L. Platon, F. Zehraoui, F.Tahi. Self-Organizing Maps with supervised layer. In Proceedings of the 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+). 2017: 161-168.
- Boucheham A, Sommard V, Zehraoui F, Boualem A, Batouche M, Bendahmane A, Israeli D, Tahi F. PLoS One 2017, 16;12(6):e0179787.

