

## PROPOSITION DE STAGE

### Master 2 ou Ingénieur (INFORMATIQUE ou BIONFORMATIQUE)

**Titre :** Modèle à base de réseaux de neurones pour l'identification des sous-types du Cancer de la Vessie

**Mots-clefs :** Réseaux de neurones, Cartes SOM, Apprentissage supervisé, Apprentissage non-supervisé, bioinformatique, prédiction de phénotypes

**Encadrants :**

Farida Zehraoui, Maître de Conférences, Equipe AROBAS, laboratoire IBISC, Université Paris-Saclay, Univ. Evry

François Radvanyi, Directeur de Recherche, Equipe Oncologie Moléculaire, Institut Curie

Constance Creux, Doctorante, Equipe AROBAS, laboratoire IBISC, Université Paris-Saclay, Univ. Evry

Fariza Tahy, Professeur, Equipe AROBAS, laboratoire IBISC, Université Paris-Saclay, Université d'Evry

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

**Durée de stage :** 6 mois

**Contact :** farida.zehraoui@univ-evry.fr

**Résumé :**

La médecine de précision, aussi appelée médecine stratifiée, consiste à utiliser principalement les caractéristiques génomiques des patients pour une prise en charge personnalisée.

Les technologies "omiques" : génomique (séquençage ADN), transcriptomique (microarrays), protéomique, ont considérablement modifié l'échelle des données et permettent de générer des quantités massives de données génomiques sur les patients. Ces données peuvent couvrir tous les mécanismes impliqués dans les variations qui se produisent dans les réseaux cellulaires qui influencent le fonctionnement des systèmes organiques chez l'homme. Elles peuvent être exploitées pour le diagnostic, le pronostic, la prédiction de traitement personnalisé du patient, etc.

L'intelligence artificielle, en particulier l'apprentissage automatique, est devenue au cours de la dernière décennie un outil prometteur pour l'aide à la médecine de précision en oncologie. L'apprentissage profond, qui est un sous-domaine de l'apprentissage automatique, jouera un rôle majeur dans l'amélioration de la précision des prédictions de susceptibilité, de récurrence et de survie au cancer.

Le cancer étant une maladie hétérogène, plusieurs sous-types peuvent être identifiés. Les traitements et les diagnostics doivent être adaptés à chaque sous-type. Dans ce projet, nous nous intéresserons à la prédiction de sous-types de d'un cancer fréquent chez l'homme, le cancer de vessie.

Le but de ce stage est de développer une méthode d'apprentissage automatique multi-sources combinant l'apprentissage supervisé et le non-supervisé, afin d'identifier les sous-types de cancer et éventuellement d'en découvrir des nouveaux, en utilisant des sources de données hétérogènes représentant différents types de données omiques associées aux patients. La méthode se basera sur des réseaux de neurones supervisés de type perceptron multi-couches (ou MLP) et des réseaux de neurones non-supervisés de type cartes auto-organisatrices (ou SOM : Self Organizing Maps) pour permettre une meilleure visualisation et interprétation des résultats.

## Bibliographie :

- V. Bourgeais, F. Zeharoui, B. Hanczar. GraphGONet: a self-explaining graph-based neural network encapsulating the Gene Ontology for phenotype prediction on gene expression. *Bioinformatics*, 2022.
- V. Bourgeais, F. Zeharoui, B. Hanczar. Deep GONet: Self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data. *BMC bioinformatics* 22, 455, 2021.
- Graw, S., Chappell, K., Washam, C.L., Gies, A.J., Bird, J.T., Robeson, M.S., & Byrum, S.D. (2020). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular omics*..
- B. Hanczar, F. Zehraoui, T. Issa, M. Arles. Biological interpretation of deep neural network for phenotype prediction based on gene expression. *BMC bioinformatics* 21 (1), 1-18, 2020.
- 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, Somnard V, Zehraoui F, Boualem A, Batouche M, Bendahmane A, Israeli D, Tahi F. *PLoS One* 2017, 16;12(6):e0179787.