

## Proposition de sujet de stage de Master 2

### *Multi-omics dimensionality reduction approaches for the study of Sepsis*

### *Approches de réduction de dimensionnalité pour traiter les données multi-omiques pour l'étude de Sepsis*

## Description

### Contexte général :

Ce sujet de Master s'inscrit dans le cadre du projet ANR RHU RECORDS (*Rapid Recognition of Corticosteroid Resistant or Sensitive Sepsis*) qui a pour objectif d'identifier et de valider des biomarqueurs de prédiction de la réponse thérapeutique aux corticoïdes dans le cadre du sepsis. Parmi ces biomarqueurs, les "omiques" ont actuellement une place particulière car elles permettent l'exploration approfondie et non biaisée, libre de toute hypothèse préalable, d'une couche (ou "layer") d'un système vivant.

Les "omiques" parmi les plus analysées en médecine sont :

- la génomique, c'est-à-dire dans le contexte de cette discussion, l'étude des variants héréditaires de l'ADN,
- la transcriptomique, ou l'étude des transcrits, qui sont les premiers produits de l'expression de l'ADN au niveau des gènes,
- la protéomique, ou l'étude des protéines, obtenues après la traduction des transcrits,
- la métabolomique, qui étudie les métabolites résultant de l'activité des protéines.

Ces omiques correspondent en fait au flux d'information génétique qui part de l'ADN et aboutit au phénotype final et contribue, avec les facteurs de l'environnement, à sa variabilité inter-individuelle. La compréhension des mécanismes de cette variabilité est essentielle car c'est elle qui détermine la susceptibilité individuelle aux maladies et la réponse des patients aux traitements de ces maladies.

La contrepartie de la puissance des méthodes omiques et du caractère massif (surtout en terme de dimensionnalité) des données qu'elles génèrent réside dans la complexité de leur analyse. Si l'analyse des omiques considérées une par une (single omics) a fait des progrès considérables sur le plan méthodologique, il peut arriver, par exemple pour des raisons de puissance statistique, que les résultats obtenus soient peu concluants.

### Motivations et objectifs du projet :

Le projet RECORDS s'appuie sur trois de ces omiques, la génomique, la transcriptomique et la métabolomique, qui sont parmi les plus robustes et de loin les plus utilisées dans l'exploration des maladies humaines. Dans ce cadre, la recherche de biomarqueurs (pas

seulement omiques) de prédiction de la réponse aux corticoïdes dans le sepsis peut se faire avec la perspective suivante (entre autres) :

On recherche des marqueurs différentiellement représentés entre les différents groupes de patients (par exemple, patients sensibles ou non au traitement par les corticoïdes). On parle d'apprentissage supervisé. Dans ce cas, on est confronté à une situation particulière qui est celle où l'on dispose d'une liste de marqueurs trop longue pour être exploitée dans la pratique médicale quotidienne. Il faut alors réduire cette liste en sélectionnant les variables les plus pertinentes qui permettront de développer des tests rapides, faits au lit du patient, pour une prise de décision thérapeutiques dans les heures suivant l'hospitalisation du patient. La réduction de dimensions est un problème classique en analyse de données et en apprentissage. Cependant, il existe des méthodes spécifiques pour les données multi-omiques, lesquelles intègrent les différentes couches. Un benchmark de ces méthodes a été publié récemment dans le contexte du cancer (Cantini et al., 2021). Ainsi, l'objectif principal de ce stage est d'étudier l'application des méthodes existantes de réduction de dimensions, dédiées aux données multi-omiques, sur les données omiques obtenues dans le cadre du RHU RECORDS, et de proposer une nouvelle méthode de réduction de dimensions en fonction des limites qui seront identifiées.

### **Méthodologie de travail :**

La méthodologie de travail qui sera suivit par le ou la candidat(e) est comme suit :

1. Étudier et appliquer les méthodes de l'état de l'art dédiées à la réduction de dimensions pour les données multi-omiques de RHU RECORDS.
2. Identifier les limites de celles-ci
3. Proposer une nouvelle méthode qui permet de pallier aux limites identifiées.
4. Résumer le travail réalisé dans un document de recherche scientifique qui sera soumis à une conférence/une revue internationale.

### **Références bibliographiques :**

1. Krassowski, Michal, et al. "State of the field in multi-omics research: From computational needs to data mining and sharing." *Frontiers in Genetics* 11 (2020).
2. Cantini L, Zakeri P, Hernandez C, Naldi A, Thieffry D, Remy E, Baudot A. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature Commun.* 2021 Jan 5;12(1):124. doi: 10.1038/s41467-020-20430-7.
3. Meng C, Zeleznik OA, Thallinger GG, Kuster B, Gholami AM, Culhane AC. Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in bioinformatics.* 2016 Jul 1;17(4):628-41.
4. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances.* 2021 Mar 29:107739.

### Profil et expertises recherchés :

Le ou la candidat(e) doit être titulaire d'une Licence en informatique ou bio-informatique. Elle/il devrait avoir :

- Compétences solides en programmation objet et système et bases de données;
- Bonne expérience en bio-informatique (souhaitable mais non obligatoire);
- Bonne expérience en statistiques;
- Bonne expérience en machine learning;
- Bonne communication orale en anglais, compétences techniques en lecture et en écriture;
- La maîtrise du français est souhaitable mais pas obligatoire.

### Durée du stage et rémunération :

- Durée du stage : 5 à 6 mois
- Rémunération selon la réglementation en vigueur sur la gratification des stages (estimée entre 550 et 600€ par mois).

### Procédure de candidature :

- Adresser aux co-encadrantes les éléments suivants :  
CV détaillé, lettre de motivation, bulletins de notes des 2 dernières années, lettres de recommandation éventuelles

**Date limite de candidature : 28 février 2022**

### Encadrements & contacts :

- Laboratoire DAVID, UVSQ, Université Paris-Saclay : Prof. Karine ZEITOUNI, [Karine.Zeitouni@uvsq.fr](mailto:Karine.Zeitouni@uvsq.fr) ; Dr. Zaineb CHELLY DAGDIA, [zaineb.chelly-dagdia@uvsq.fr](mailto:zaineb.chelly-dagdia@uvsq.fr), co-encadrantes.
- En collaboration avec Prof. Henri-Jean Garchon, [henri-jean.garchon@uvsq.fr](mailto:henri-jean.garchon@uvsq.fr), Prof. Stanislas Grassin Delye, [stanislas.grassin-delye@uvsq.fr](mailto:stanislas.grassin-delye@uvsq.fr),

### Laboratoires d'accueil :

- [DAVID/équipe ADAM](#), Campus de sciences à Versailles, Université de Versailles St-Quentin / Université Paris-Saclay. Lieu du stage.
- [Laboratoire IIC, U1173](#), UFR Simone Veil, Saint-Quentin-en-Yvelines