

Stage M2 : Méthodes computationnelles innovantes pour l'analyse haut-débit des données métabolomiques et la découverte de biomarqueurs en santé

Contexte du projet

La métabolomique par spectrométrie de masse (MS) est une approche de pointe pour l'exploration chimique du vivant et la découverte de biomarqueurs en santé : elle a pour objectif de détecter, quantifier et identifier les milliers de petites molécules présentes dans les échantillons biologiques, et de développer des modèles prédictifs de la survenue d'une maladie ou de la réponse à un traitement.

Aussi, l'analyse computationnelle haut-débit des données MS constitue un enjeu majeur. Celles-ci se présentent en 3D (masse, temps de rétention, intensité) sous la forme de pics correspondant aux molécules détectées, et sont à la fois volumineuses ($\sim 10^2$ Mo par échantillon) et complexes. L'équipe « science des données » développe depuis plus de dix ans des méthodes et des logiciels innovants en traitement du signal et en machine learning pour les données MS.

Sujet proposé

Pour approfondir l'identification chimique en routine des composés, des approches technologiques prometteuses (dites MS/MS ou « de fragmentation ») se sont développées récemment. L'objet du stage porte sur l'analyse computationnelle des données combinées MS et MS/MS pour la découverte de biomarqueurs en recherche clinique.

La première partie du stage portera sur le traitement du signal de ces données pour extraire les spectres de fragmentation de qualité optimale, notamment grâce aux méthodes de déconvolution 2D (séparation de sources positives, régression parcimonieuse, clustering hiérarchique) pour éliminer les contaminations diverses (bruit, molécules co-fragmentées, etc.).

La seconde partie sera consacrée à l'analyse des spectres MS/MS par des approches de fouilles de données (e.g. recherche de motifs de pics fréquents) et de machine learning/deep learning (e.g. prédiction de fonction ou de classe chimique du composé à partir des bases de données spectrales ou moléculaires publiques).

Profil recherché

Nous recherchons un(e) étudiant(e) en Master 2 avec une solide formation en mathématiques appliquées (traitement du signal), et en informatique (logiciels R et C), et motivé(e) par les applications multidisciplinaires (chimie, biologie, clinique). Le master pourra se prolonger par une thèse.

Lieu du stage

Département Médicaments et Technologies pour la Santé (MTS/SPI)
CEA Saclay, F-91191 Gif sur Yvette, France

Durée du stage : 6 mois à partir de mars 2022

Date de clôture des candidatures : 01/12/2021

Contact

Etienne Thévenot (etienne.thevenot@cea.fr)

Equipe Sciences des Données

Web: <https://scidophenia.github.io/>