

Utilisation des images RMN 2D pour la quantification en métabolomique

Proposition de stage niveau M2 (printemps 2023)

Unité LBE, INRAE de Narbonne

Pour postuler envoyer CV, lettre de motivation et notes de Master à gaelle.lefort@inrae.fr ; remi.servien@inrae.fr et marie.tremblay-franco@inrae.fr

Contexte

Les branches de la science connues sous le nom de **omiques** (génomique, transcriptomique, métabolomique ...) visent à la compréhension d'un organisme biologique. Elles offrent un cadre de travail récent et ont en commun la production de large jeux de données, menant à des problèmes modernes en **statistique et analyse de données**. Nous nous intéressons ici à la **métabolomique**, qui étudie l'ensemble des métabolites (sucres, acides aminés ...) présents dans un organisme d'intérêt. La métabolomique peut permettre de mieux comprendre certains phénomènes biologiques (comme les différences de maturité entre certaines races de porc à la naissance (Lefort et al., 2020)) mais l'exploitation des données générées reste délicate. En effet, **l'identification et la quantification** automatique des métabolites dans les spectres produits reste perfectible. Le projet ANR **PANORAMICS** vise à rendre cette procédure plus fiable en utilisant de manière complexe et automatique des images 2D également générées lors de l'obtention des données.

Mission

La première étape sera de comparer différentes méthodes automatiques de **traitement d'images** afin de réaliser de manière automatique l'identification des pics sur les images métabolomiques. Certaines méthodes (Hahsler et al., 2019 ; Ross et al., 2022) ont déjà été utilisées par les encadrantes de ce stage afin de réaliser des pré-tests mais les premiers résultats obtenus doivent être confirmés et comparés à des outils disponibles en ligne (Li et al., 2022). Nous disposons pour cela de données obtenues sur des échantillons synthétiques dont la composition est connue. La seconde étape sera de combiner cette détection avec **l'algorithme BARSA**, qui permet l'identification de métabolites dans les images 2D et qui a été développé par une co-encadrante du stage, afin de le rendre automatique. La dernière étape sera de déterminer comment combiner dans le modèle statistique l'information extraite de ces images avec celles extraites des spectres, par exemple par le **package R ASICS** (Lefort et al., 2019). Ce travail, qui sera réalisé sous le logiciel R, pourra donner lieu à une **publication scientifique et/ou un package R**. Il pourra éventuellement être suivi par un CDD d'un an sur un sujet connexe.

Profil recherché

- Master 2 ou dernière année d'école d'ingénieur en statistique/mathématiques appliqués ;
- Maîtrise d'un langage de programmation scientifique (R, Matlab, Python) ;
- Bonnes connaissances en statistiques multivariées ;
- Aucune connaissance préalable en chimie ou en biologie n'est nécessaire.

Conditions du stage

durée 4 à 6 mois

localisation Unité LBE, INRAE de Narbonne

rémunération taux légal (environ 560 euros)

encadrement Gaëlle Lefort, Rémi Servien, Marie Tremblay-Franco

Contact : gaelle.lefort@inrae.fr ; remi.servien@inrae.fr et marie.tremblay-franco@inrae.fr

Bibliographie

1. G. Lefort, R. Servien, H. Quesnel, Y. Billon, L. Canario, N. Iannuccelli, C. Canlet, A. Paris ,N. Vialaneix and L. Liaubet. The maturity in fetal pigs using a multi-fluid metabolomic approach. *Scientific Reports* (2020), **10**, 19912. [Pdf sur BiorXiv](#). [Lien éditeur \(open access\)](#).
2. G. Ross, D. Markwick, K. Mulder and G. Sighinolfi. R package dirichletProcess. Available on CRAN. [Lien](#).

3. M. Hahsler, M. Piekenbrock, and D. Doran, D. dbscan : Fast Density-Based Clustering with R, *Journal of Statistical Software* (2019), **91**(1), 1-30. [Lien éditeur](#).
4. D. Li, A. Leggett, L. Brusweiler-Li, and R. Brusweiler. COLMARq : A Web Server for 2D NMR Peak Picking and Quantitative Comparative Analysis of Cohorts of Metabolomics Samples, *Analytical Chemistry* (2022), **94**(24), 8674-8682. [Lien éditeur](#).
5. G. Lefort, L. Liaubet, C. Canlet, P. Tardivel, M.-C. Pere, H. Quesnel, A. Paris, N. Iannucelli, N. Vialaneix and R. Servien. ASICS : an R package for a whole analysis workflow of 1D ^1H NMR spectra. *Bioinformatics* (2019), **35**(21), 4356-4363. [Pdf sur BiorXiv](#). [Lien éditeur](#). [R package sur Bioconductor](#).