

Analyse de données multi-blocs et trois-voies pour une meilleure compréhension des liens entre la composition du microbiote intestinal et la maturation du système immunitaire.

Mots-clés : Données multi-blocs, Données trois-voies, Sparsité, Microbiote, Allergie infantile.

Contexte :

Les allergies touchent 30-40% de la population mondiale. Parmi elles, la dermatite atopique est la plus fréquente et précoce. Ces allergies sont associées à un déséquilibre de la flore intestinale et à un dysfonctionnement du système immunitaire chez les nouveaux-nés. En induisant une flore optimale, les prébiotiques donnés tôt dans la vie pourraient prévenir des allergies. Le projet ANR CIMMAP (Characterising the effect of maternal prebiotic supplementation on perinatal Immune system maturation, Gut Microbiota and breast Milk compositions for Allergy Prevention in high-risk children) est un projet de recherche ANR collaboratif (voir <https://www6.angers-nantes.inrae.fr/bia/Equipes-de-recherche/Allergie-ALL/Actualites/ANR-CIMMAP>).

En particulier, CIMMAP explore la période périnatale chez l'enfant à risque d'allergie en se focalisant sur la mise en place du système immunitaire et du microbiote et les effets des prébiotiques sur le système immunitaire et sur le microbiote. Les données considérées dans CIMMAP sont issues dans un premier temps d'un modèle murin et dans un second temps d'un modèle humain, sur la base d'une cohorte sélectionnée à partir de l'étude PREGRALL (<https://www.chu-nantes.fr/pregrall-participez-a-notre-etude-et-aidez-nous-a-prevenir-les-allergies-de-l-enfant>).

Objectif du stage :

Des travaux antérieurs sur le modèle murin ont permis de mesurer les effets d'une supplémentation en prébiotiques d'une part sur la composition du microbiote et d'autre part sur les caractéristiques du système immunitaire. L'objectif de ce stage est d'investiguer une approche holistique visant à mettre en relation ces deux blocs de données, en prenant également en compte une dimension longitudinale (plusieurs points de prélèvement effectués au cours de l'étude). Cette approche holistique repose sur un traitement de données multi-blocs (plusieurs tableaux de données, appariés sur les lignes c'est-à-dire les individus, sont analysés simultanément pour révéler la force des liens entre ces tableaux) voire un traitement de données trois voies (les tableaux de données étant cette fois appariés à la fois sur les individus et sur les variables).

En particulier deux verrous statistiques seront appréhendés.

Dimensionnalité des tableaux et sparsité : dans le cadre des données de microbiote, le tableau de données présente un grand nombre de variables, avec existence d'une forte multicollinéarité. Un premier enjeu est donc de pouvoir gérer la dimensionnalité de ce tableau

au regard des données d'immunologie, en particulier de manière à mesurer le lien entre ces deux tableaux et identifier un petit nombre de marqueurs du microbiote fortement liés aux caractéristiques du système immunitaire. Pour répondre à ce premier enjeu, le stage visera à étudier et mettre en œuvre des méthodes multi-blocs sparse dans un contexte non supervisé (ComDim, MB-PCA, RGCCA) et supervisé (P-ComDim, MB-PLS, MB-WCov).

Introduction de la dimension longitudinale : dans un second temps, une approche longitudinale sera investiguée pour prendre en compte simultanément les tableaux de données du microbiote prélevés à différents temps. Il s'agira en particulier d'identifier des trajectoires communes pour des groupes de marqueurs par des méthodes de décomposition tensorielle.

Profil souhaité du candidat

Niveau Bac +5, Master II ou ingénieur, avec une formation en statistique ou biostatistique.

Intérêt pour le traitement statistique de données biologiques et de santé.

Maîtrise des méthodes d'analyse multivariée dans le cadre supervisé et non supervisé.

Maîtrise du langage R.

Lieu, durée et rémunération :

Le stage se déroulera à Oniris, campus de la Géraudière, 44032 NANTES, au sein de l'unité Statistique, Sensométrie et Chimiométrie. Cette unité, sous contrat INRAE n° 1381, est composée de 5 enseignants-chercheurs et de deux ingénieurs de recherche.

Le stage est prévu pour une durée de 6 mois et commencera au plus tard en mars 2022.

La rémunération est celle d'une gratification de stage (entre 550 et 600 euros par mois).

Contact :

Un CV et une lettre de motivation sont à envoyer à :

Véronique Cariou (veronique.cariou@oniris-nantes.fr)

El Mostafa Qannari (elmostafa.qannari@oniris-nantes.fr)

Bibliographie :

Karaman, İ., Nørskov, N. P., Yde, C. C., Hedemann, M. S., Knudsen, K. E. B., & Kohler, A. (2015). Sparse multi-block PLSR for biomarker discovery when integrating data from LC-MS and NMR metabolomics. *Metabolomics*, 11(2), 367-379.

Lê Cao, K. A., Rossouw, D., Robert-Granié, C., & Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Statistical applications in genetics and molecular biology*, 7(1).

Lorenzo, H., Saracco, J., & Thiébaud, R. (2019). Supervised learning for multi-block incomplete data. *arXiv preprint arXiv:1901.04380*, 2019

Mangamana, E. T., Cariou, V., Vigneau, E., Kakaï, R. L. G., & Qannari, E. M. (2019). Unsupervised multiblock data analysis: A unified approach and extensions. *Chemometrics and Intelligent Laboratory Systems*, 194, 103856.

Mangamana, E. T., Kakaï, R. G., & Qannari, E. M. (2021). A general strategy for setting up supervised methods of multiblock data analysis. *Chemometrics and Intelligent Laboratory Systems*, 217, 104388.