

Post-doctorat en sciences des données biologiques : Méthodes innovantes d'intégration de données multi-omiques appliquées à la prédiction de l'allergie alimentaire

Contexte

Les approches globales de type « omiques » (p. ex. la métabolomique) sont d'un grand intérêt pour la compréhension du métabolisme humain et la prédiction des maladies. L'analyse de tels jeux de données (caractérisés par un nombre de variables supérieur à celui des échantillons, et par la présence de colinéarité entre les variables) nécessite des méthodes multivariées spécifiques pour l'analyse non-supervisée comme pour la prédiction (Thévenot et al., 2015 ; Rinaudo et al., 2016) que notre équipe développe depuis de nombreuses années au sein du Laboratoire Sciences des Données et Décision (CEA).

Aujourd'hui, la combinaison d'analyses omiques complémentaires (p. ex. la métabolomique et la lipidomique) apparaît comme une approche prometteuse pour élargir la liste des biomarqueurs et augmenter les performances de prédiction. De nouvelles méthodes statistiques sont donc nécessaires pour modéliser de tels jeux de données multi-tableaux.

Pour comprendre l'impact de l'environnement maternel sur la composition précoce du lait maternel et le développement d'allergies alimentaires, 300 échantillons de lait de la cohorte mère-enfant EDEN ont été analysés par métabolomique, lipidomique, glycomique et par les mesures ciblées de marqueurs immuns, dans le cadre d'un partenariat entre le Laboratoire d'Etude du Métabolisme des Médicaments (CEA) et le Laboratoire d'Immuno-Allergie Alimentaire (INRA).

Projet

L'objectif du projet est de développer de nouvelles méthodes biostatistiques pour intégrer les cinq jeux de données ainsi que les covariables cliniques, et construire des modèles prédictifs des allergies alimentaires robustes et précis. Des approches linéaires (analyse de données multi-blocs, réseaux de corrélation partielle) et non linéaires seront utilisées, en plus de l'analyse de réseaux biochimiques (pour inclure des informations biologiques et chimiques supplémentaires).

Les défis comprendront notamment la sélection d'une signature multi-omique restreinte, le traitement des effets confondants, des temps de collecte distincts des échantillons et de l'hétérogénéité de la classe « allergie ».

Les méthodes seront mises en œuvre dans le langage R.

Profil

Les candidats intéressés doivent être titulaires d'un doctorat en statistiques appliquées (biostatistique, analyse des données, machine learning, sélection de variables, analyse de graphe) et être motivés par des applications multidisciplinaires (chimie et biologie).

Contact

Etienne Thévenot

CEA, LIST, Centre de Saclay, Laboratoire Sciences des Données et Décision

Email: etienne.thevenot@cea.fr

Web: <https://frama.link/etienne-thevenot>