Evaluation de méthodes statistiques pour l'intégration de données métabolomiques, cliniques et alimentaires

Marion Brandolini-Bunlon*†[1], Mélanie Petera[1], Stéphanie Monnerie[2], Charlotte Joly[1], José Morais[3], Hélène Payette[5,4], Pierrette Gaudreau[6,7], Blandine Comte[2], Julien Boccard[8], and Estelle Pujos-Guillot[1,2]

[1]Université Clermont Auvergne, INRA, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont – Institut national de la recherche agronomique (INRA) : UMR1019 – F-63000 Clermont-Ferrand, France

[2]Université Clermont Auvergne, INRA, UNH – Institut national de la recherche agronomique (INRA) : UMR1019 – F-63000 Clermont-Ferrand, France

[3]Division of Geriatric Medicine, McGill University – Montreal, Canada

[5]Faculté de médecine et des sciences de la santé, Université de Sherbrooke – Sherbrooke, Canada

[4]Centre de recherche sur le vieillissement, Centre intégré universitaire de santé et de services sociaux de Sherbrooke – Sherbrooke, Canada

[6]Département de médecine, Université de Montréal – Montréal, Canada

[7]Centre de Recherche du Centre hospitalier de l'Université de Montréal – Montréal, Canada

[8]School of Pharmaceutical Sciences, University of Lausanne, University of Geneva – Geneva, Suisse

Résumé

Du fait de la prévalence croissante des maladies métaboliques chroniques, la compréhension de leurs mécanismes et du rôle de la nutrition dans leur développement et leur prévention est aujourd'hui un enjeu majeur. Leur complexité (maladies de systèmes) requiert une approche populationnelle, intégrée et multifactorielle. L'objectif de ce travail est de proposer une stratégie d'intégration de données afin de construire des modèles multidimensionnels, descriptifs ou discriminants de pathologies pour la découverte de biomarqueurs, et de contribuer à une meilleure compréhension des processus physiopathologiques impliqués.
Pour cela, cette étude s'est basée sur des données et échantillons biologiques de sujets de la cohorte NuAge du Réseau Québécois de Recherche sur le vieillissement : 123 hommes (68-82 ans) ont été sélectionnés pour une étude cas/témoins (61/62 sujets) du syndrome métabolique. Ils ont été suivis annuellement avec des prélèvements biologiques, la mesure de paramètres cliniques et le recueil de données par questionnaires. Des analyses métabolomiques ont été réalisées par spectrométrie de masse sur des échantillons de sérum collectés en 2003-2005 (T1) et 3 ans plus tard (T4). Trois blocs de données ont ainsi été identifiés : métabolomiques (863 ions), cliniques (29 variables) et alimentaires (41 variables), par temps et par groupe de sujets. Ils ont été analysés avec le logiciel R, simultanément ou deux à deux, après normalisation UV ou Pareto, par des méthodes d'analyses multiblocs, à des fins descriptives (ACP multiblocs et Analyse Factorielle Multiple (AFM)) ou discriminantes (PLS

*Intervenant

†Auteur correspondant: marion.brandolini-bunlon@inra.fr

multiblocs). Les résultats sont comparés sur la base de critères statistiques (pourcentage de variance expliquée, Q^2, R^2, matrice de confusion...) et de leur adéquation aux objectifs biologiques : discrimination cas/témoins, identification de marqueurs, découverte de sous-groupes, adéquation aux connaissances du domaine...
Ce travail a permis d'établir une stratégie d'intégration de données pour mettre en œuvre une approche multidimensionnelle d'étude des pathologies systémiques.

Mots-Clés: intégration de données, ACP multiblocs, PLS multiblocs, AFM, maladies métaboliques chroniques, nutrition, spectrométrie de masse, biomarqueurs