

Stage Master II

Intégration statistique de données « omiques » : application à l'étude de l'effet des mycotoxines

CONTEXTE

Le stage proposé se déroulera au sein du Laboratoire de Toxicologie Alimentaire de l'INRA de Toulouse, qui s'intéresse aux effets des contaminants toxiques des aliments sur la santé de l'Homme et de l'animal, en utilisant entre autres des approches globales de biologie intégrative. En particulier, le laboratoire s'intéresse aux mycotoxines, métabolites secondaires fongiques toxiques produits par différentes moisissures qui peuvent contaminer les céréales en Europe. De par leur régime riche en céréales, les porcs sont particulièrement exposés à ces mycotoxines. De nombreux effets délétères dus à la présence de ces mycotoxines sur la santé des animaux ont été montrés. L'identification de biomarqueurs d'exposition et d'effet des mycotoxines apparaît ainsi importante pour le suivi de la santé des animaux, ainsi que pour une meilleure compréhension du mécanisme d'action de ces mycotoxines.

La recherche de ces biomarqueurs a grandement bénéficié de l'avènement de la génomique et des avancées des technologies à haut-débit qui permettent d'acquérir un ensemble de données à l'échelle du génome (biomolécules – ADN, ARNm, protéines et métabolites – correspondant aux différents niveaux fonctionnels d'un organisme et reliés par des réactions), connus sous le nom de données 'omiques.

Dans de nombreuses études, plusieurs jeux de données « omiques » sont désormais générés à partir des mêmes échantillons. Ces jeux de données contiennent des informations complémentaires sur le système biologique étudié et sont souvent traités de façon indépendante. L'analyse conjointe (appelée intégration ou fusion) de ces données pourrait permettre d'aborder le système biologique dans sa globalité pour mieux en comprendre le fonctionnement dans toute sa complexité.

L'intégration de données 'omiques est confrontée à plusieurs problèmes méthodologiques (dimensionnalité et hétérogénéité des jeux de données, entre autres). De nombreuses méthodes d'intégration ont été proposées dans la littérature, telles que des variantes pénalisées de l'Analyse Canonique des Corrélations, les cartes de Kohonen ou des méthodes basées sur la régression des moindres carrés partiels.

OBJECTIF

L'objectif du stage est de proposer une (ou plusieurs) méthode(s) adaptée(s) à l'intégration de plusieurs jeux de données omiques (transcriptomique, protéomique, métabolomique) afin d'identifier des biomarqueurs d'effet suite à une exposition aux mycotoxines, et les voies métaboliques perturbées par ces contaminants.

Dans ce but les tâches suivantes seront mises en œuvre :

- Inventorier et évaluer les différentes méthodes basées sur la régression des moindres carrés partiels disponibles dans la littérature
- Proposer une méthode adaptée à nos jeux de données (sur la base de données simulées)
- Réaliser l'analyse de données avec la méthode sélectionnée

RÉFÉRENCES BIBLIOGRAPHIQUES

- M. Y. Hirai, M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara and K. Saito (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. PNAS 101(27): 10205-10210.
- J. Boccard, D.N. Rutledge (2013). A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Analytica Chimica Acta* 769: 30-39.
- R. Cavill, D. Jennen, J. Kleinjans, J. Jan Briedé (2016). Transcriptomic and metabolomic data integration. *Briefings in Bioinformatics* 17(5):891–901.
- C. Meng, O.A. Zeleznik, G.G. Thallinger, B. Kuster, A.M. Gholami, A.C. Culhane (2016).

Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief Bioinform.* 17(4): 628–641.

- I.S.L. Zeng, T. Lumley (2018). Review of Statistical Learning Methods in Integrated Omics Studies (An Integrated Information Science). *Bioinform Biol Insights* 12.

- O.P. Günther, H. Shin, R.T. Ng, W.R. McMaster, B.M. McManus, P.A. Keown, S.J. Tebbutt, K-A. Lê Cao (2014). Novel Multivariate Methods for Integration of Genomics and Proteomics Data: Applications in a Kidney Transplant Rejection Study. *OMICS: A Journal of Integrative Biology* 18(11): 682-695.

Adresse

UMR TOXALIM - INRA

180 chemin de Tournefeuille

BP 93173

31027 TOULOUSE 3

France

Contact

Marie Tremblay Franco

marie.tremblay-franco@inra.fr

deadline : 10/11/2019