

# Compréhension du lien microbiote-nutrition par les statistiques multivariées et le machine learning

Nous sommes en permanence en interaction avec une myriade de micro-organismes qui vivent dans nos intestins. L'ensemble de ces micro-organismes est appelé microbiote intestinal.

La composition du microbiote intestinal humain est associée à de nombreuses maladies chroniques comme l'obésité (Cotillard et al., 2013; Le Chatelier et al., 2013), le diabète de type 2 (Qin et al., 2012) ou la cirrhose (Qin et al., 2014) pour lesquelles nous ne disposons pas de traitement satisfaisant. Nous espérons qu'une meilleure compréhension du microbiote intestinal mènera à la création de traitements innovants ou de méthodes de prévention efficaces.

MetaGenoPolis a développé un pipeline complet qui permet de mesurer l'abondance des micro-organismes présent dans l'intestin à partir d'échantillons fécaux. Pour chaque cohorte, on dispose d'une matrice échantillons versus micro-organismes. Classiquement, les cohortes sont composées de quelques centaines d'individus alors que le microbiote intestinal comporte environ un millier d'espèces.

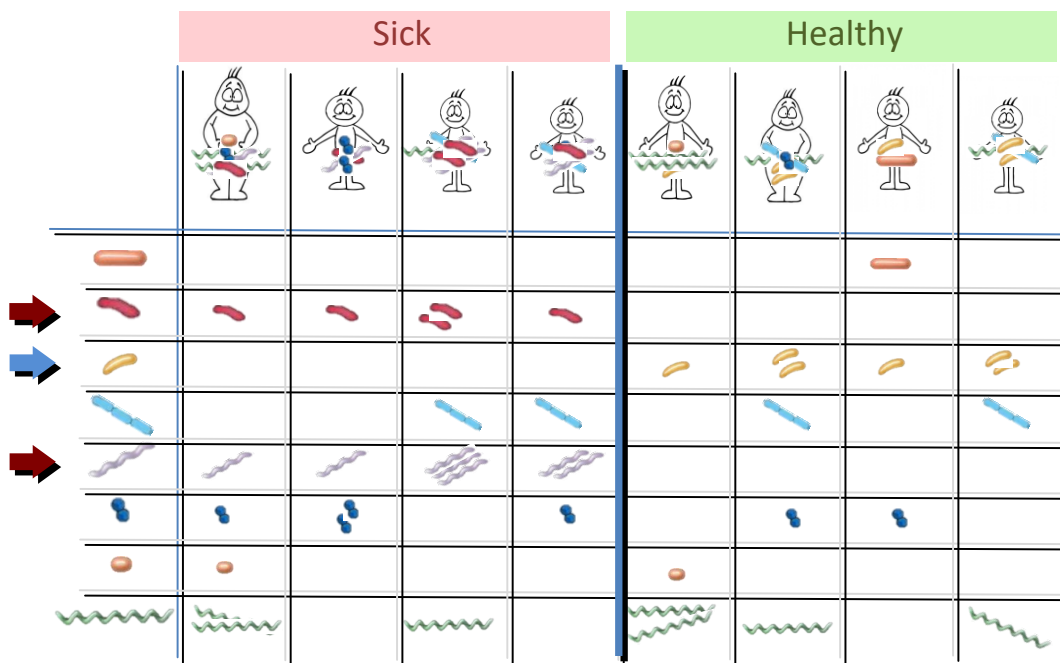


Figure 1: Changement de composition du microbiote en fonction du phénotype des individus

La tâche des analystes de MetaGenoPolis est de donner du sens à ces matrices en mettant en lien l'abondance des micro-organismes avec le phénotype des individus. Par exemple, le package [MetaOMiner](#) nous permet de détecter les espèces significativement enrichies chez le malade.

Ce type de données présente cependant un challenge important pour le statisticien. Un premier facteur de complexité est la haute dimensionnalité des données. En effet le nombre de variables prises en compte est en général largement supérieur aux nombres d'échantillons séquencés. Ce type de problème requiert une adaptation des méthodes classiques ou des méthodes particulières:

- Réduction de la dimensionnalité pour synthétiser les données (Analyse en Composante Principale, clustering, ...)
- Régularisation pour éviter le sur-apprentissage (régression régularisée, Lasso)

Dans notre contexte, une seconde source de complexité provient l'hétérogénéité des données. Nous avons d'une part la composition du microbiote et d'autre part les quantités d'aliments ingérés. Nous souhaitons décrire au mieux les corrélations entre ces deux ensembles de données pour avoir une vue d'ensemble de l'impact des aliments ingérés sur le microbiote.

Pour répondre à cette question, on appliquera une méthode qui est à fois une réduction de dimensionnalité et une analyse de la corrélation : **L'analyse canonique des corrélations**. Le stage consistera à :

1. appliquer cette méthode sur plusieurs cohortes,
2. à interpréter les résultats en termes de nutrition et de microbiologie (avec l'appui de l'encadrante),
3. à vérifier la cohérence des conclusions entre ces différentes cohortes.

Le stage laissera la latitude à ceux qui le souhaitent de proposer des méthodes équivalentes pour répondre à la question.

## PROFIL :

Vous êtes un étudiant en Machine learning ou data science qui souhaite appliquer ses connaissances dans un domaine intellectuellement fascinant et ayant un impact sur la santé humaine.

Vous êtes un biostatisticien (bio-informaticien) qui souhaite approfondir ses connaissances en statistiques multivariées et en machine learning.

## CONNAISSANCES TECHNIQUES SOUHAITEES OU SOUHAITABLES:

- Bonne connaissance du langage R ou du langage python (dans ce cas il est recommandé de connaître le package scikit-learn et pandas)
- Connaissances de bases des statistique multivariées (Analyses en composantes principales, clustering)

Contact : [hanna.julienne@inra.fr](mailto:hanna.julienne@inra.fr)

## BIBLIOGRAPHIE:

Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500(7464):541-546. doi:10.1038/nature12506.

Cotillard A, Kennedy SP, Kong LC, et al. Dietary intervention impact on gut microbial gene richness. *Nature*. 2013;500(7464):585-588. doi:10.1038/nature12480.

Qin N, Yang F, Li A, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature*. 2014;513(7516):59-64. doi:10.1038/nature13568.

## RESSOURCES SUR LE MICROBIOTE:

- <http://www.gutmicrobiotaforhealth.com/fr/accueil/>
- <https://fr.coursera.org/learn/microbiome>
- <http://martinblaser.com/>
- Le Charme discret de l'intestin: Tout sur un organe mal aimé de Giulia Enders