

Titre : Développement d'une approche d'intégration de données multi-omiques pour expériences multi-groupes

Le développement des approches haut débit en biologie permet maintenant la production massive de données omiques pour des contextes applicatifs variés. Ces données sont fréquemment obtenues sur les mêmes individus à divers niveaux de l'échelle du vivant (transcriptomique, métabolomique, protéomique, lipidomique, métagénomique, ...), sous des formats très variés (données de comptage, spectres, images, ...) qui ne sont pas toujours directement interprétables d'un point de vue biologique, sont de très grande dimension (de nombreuses caractéristiques sont mesurées simultanément) et ont été obtenues pour un nombre d'échantillons qui reste modeste en comparaison du nombre de mesures effectuées. Il s'agit alors de les mettre en relation entre elles et avec les informations cliniques et le plan d'expérience complexe dans lesquelles elles ont été produites.

Si certaines analyses statistiques, comme l'analyse différentielle des diverses mesures relative à une donnée omique en relation avec des groupes d'individus (contrôles / traités par exemple) est maintenant bien balisée pour la plupart des types d'omiques, les besoins en méthodes d'intégration de données, c'est à dire en méthode capable d'extraire de l'information en combinant les vues provenant de plusieurs omiques, sont en pleine expansion et sont un sujet de recherche actif.

Pierre Fabre Cosmétique est engagé dans de multiples projets dans lesquels ce type de problèmes se pose et où des données omiques multiples ont été acquises. Les questions relatives à la combinaison de données omiques y sont posées soit sous la forme d'une problématique d'association (quels sont les éléments moléculaires – métabolites, bactéries, ... – que l'on retrouve en association dans les échantillons entre deux types d'omiques), soit sous la forme de la recherche de biomarqueurs (quels sont les éléments descripteurs d'un état phénotypique mesuré au travers de données cliniques multiples). Elles sont également systématiquement associées à un plan d'expérience dans lequel les échantillons sont structurés en groupes qu'il faut prendre en compte pour répondre à la question biologique ou clinique sous-jacente.

L'objectif de la thèse sera positionné sur le développement méthodologique pour l'intégration de données avec une implémentation et application aux données du projet. De manière plus précise, il s'agira de développer une approche à noyau (et d'un outil associé interactif et intuitif) pour l'intégration d'une paire de données omiques dans un cadre exploratoire et de problématiques d'association. L'approche développée durant cette thèse se veut flexible (adaptable à des types de données très variés), interprétable (capable d'identifier les variables importantes) et capable de prendre en compte le protocole expérimental et les groupes d'échantillons définis a priori pour proposer des interprétations communes et spécifiques de la question initiale au regard de ces groupes. L'approche développée sera déclinée pour l'interprétation biologique dans les divers projets de Pierre Fabre Cosmétique inclus dans la thèse.