

Analyse statistique d'une protéine impliquée dans les problèmes d'obésité en relation avec l'expression des gènes

Alyssa Imbert, Nathalie Viguerie & Nathalie Villa-Vialaneix
Unité MIAT, INRA de Toulouse & Laboratoire I2MC, INSERM Toulouse

Proposition de stage niveau M2

Contexte

L'obésité est caractérisée par un excès de tissu adipeux qui s'accompagne à long terme de complications métaboliques et cardiovasculaires dont les mécanismes d'installation sont encore à clarifier. Longtemps considéré comme un tissu de réserve énergétique, le tissu adipeux sécrète aussi des facteurs variés dont le nombre ne cesse de croître. Une meilleure connaissance des facteurs sécrétés par les dépôts adipeux est nécessaire pour mieux comprendre les complications métaboliques associées à l'obésité, notamment le diabète. Le Laboratoire de Recherche sur les Obésités a mis en évidence une nouvelle protéine produite par le tissu adipeux et qui n'a pas encore été étudiée dans le domaine de l'obésité et du diabète.

Nous disposons d'une vaste gamme d'indicateurs biologiques (données cliniques ou d'expression des gènes dans le tissu adipeux) chez des individus en surpoids ou obèses avant et après régime hypocalorique. L'objectif est ici de rechercher où se situe cette nouvelle protéine parmi les variables, paramètres plasmatiques ou expression de gènes (ARNm), expliquant l'amélioration de la sensibilité à l'insuline pendant le régime hypocalorique et dans quelle mesure cet effet est indépendant de la perte de poids.

Travail à effectuer

Le travail à effectuer sera dans le champ statistique de la fouille de données et de l'apprentissage statistique. De manière plus précise, le stagiaire devra utiliser des méthodes :

- d'inférence statistique (tests, ANOVA, ...);
- d'analyses factorielles (ACP, PLS, ...);
- d'apprentissage (forêts aléatoires, SVM, ...).

Toutes les analyses seront menées sous R.

Une bonne connaissance de ces méthodes est donc requise même si l'apprentissage de certaines d'entre elles pourra être réalisé durant le stage. Une très bonne connaissance de R est indispensable.

Par ailleurs, les données à traiter comportent divers problèmes que le stagiaire devra prendre en charge. En particulier, le nombre de valeurs manquantes est important et il conviendra de les traiter avec des méthodes d'imputation dont il faudra vérifier l'impact.

Conditions du stage

durée 4 à 6 mois

localisation Unité MIAT, INRA de Toulouse (Castanet-Tolosan)

rémunération taux légal

encadrement Alyssa Imbert, Nathalie Viguerie et Nathalie Villa-Vialaneix

Contact : envoyer CV et lettre de motivation aux **trois** adresses e-mails : alyssa.imbert@inra.fr, nathalie.viguerie@inserm.fr et nathalie.villa-vialaneix@inra.fr

Profil recherché

- niveau équivalent Master 2 en mathématiques appliquées;
- très bonne connaissance du logiciel de calcul scientifique R;

- une spécialisation en biostatistique ou des connaissances sur l'analyse de données biologiques serait un plus mais n'est pas indispensable : une formation basique aux problématiques dans ce domaine est prévue en début de stage.