

---

## Science statistique pour la génétique et la génomique

Présidente: Karen Kopciuk (Cancer Epidemiology and Prevention Research, AHS/University of Calgary)

---

---

**SHOFIQL ISLAM**, McMaster University

*Comparer la performance des composantes principales linéaires et non-linéaires dans le contexte de l'intégration des données génomiques de grande dimension*

Nous intégrons phénotype, expression des gènes et données d'expression miARN pour classer la tumeur ou la mort de patients atteint du cancer du poumon. Nous appliquons une analyse en composantes principales (ACP) linéaire et de noyau afin de réduire la dimension et d'intégrer ensuite trois domaines en se basant sur plusieurs approches non linéaires, y compris la régression logistique. Nous avons également effectué une vaste simulation pour comparer les performances. Nous observons qu'un nombre réduit de composantes principales de noyau renferme plus d'information sur un processus génétique, mais a échoué, dans ce cas, à montrer une meilleure capacité prédictive. À cette fin, il semble suffisant de réduire la dimension avec une ACP linéaire et un modèle de régression logistique.

---

**GUN HO JANG**, Ontario Institute for Cancer Research

*Une méthode de recalage du nombre de copies utilisant l'intégration objet pour séquencer les données de prochaine génération*

Les mutations ponctuelles et les modifications du nombre de copies sont des caractéristiques répandues des tumeurs. Le séquençage complet du génome est l'une des technologies les plus populaires pour les détecter. Plusieurs méthodes ont été développées pour désigner la modification du nombre de copies en utilisant les comptes lus et les fréquences d'allèles non-référence. Les différences entre la méthode de normalisation et les modèles statistiques présumés aboutissent à des résultats différents. Une méthode de recalage du nombre de copies utilisant l'intégration objet est proposée afin d'évaluer et d'améliorer les appels du nombre de copies. La performance de la méthode proposée est présentée et comparée à plusieurs appelants du nombre de copies, y compris HMMcopy, TitanCNA et Control-FREEC.

---

**LI LI**, Dalhousie University

*Nouvelle approche pour détecter la recombinaison génétique dans la phylogénèse*

La recombinaison génétique est un processus dans lequel les parties de différents gènes sont combinées pour former un nouveau gène. La recombinaison a de nombreux effets sur l'analyse phylogénétique de séquences d'ADN, et doit être identifiée correctement. Nous proposons un nouvel algorithme permettant de détecter la recombinaison en fonction de la matrice locale du logarithme de la vraisemblance. Cette méthode informatique est efficace et montre de grandes capacités à détecter des recombinaisons. Le rendement de cette méthode est évalué au moyen d'exemples de données simulées et réelles. Les résultats sont comparés avec ceux d'autres méthodes.

---

**KUN LIANG**, University of Waterloo

*Modèle d'arbre de Markov caché pour tester les nombreuses hypothèses de l'ontologie des gènes sur des ensembles de gènes*

Les tests sur des catégories de gènes prédéfinies sont devenus une pratique commune pour les scientifiques qui analysent les données à débit élevé provenant de transcriptomes. Une façon systématique de tester les catégories de gènes amène à tester des centaines d'hypothèses nulles qui correspondent à des nœuds dans un graphe acyclique orienté. Les relations entre les catégories de gènes entraînent des restrictions logiques entre les hypothèses nulles correspondantes. Selon un modèle d'arbre de Markov caché, nous développons une méthode informatique efficace pour incorporer l'information de dépendance parmi les hypothèses nulles. Notre méthode fournit des résultats plus puissants que ceux provenant des méthodes existantes qui respectent les restrictions logiques.

---

**FARHAD SHOKOOHI**, McGill University

*Analyse de données de méthylation de l'ADN : une comparaison des outils d'analyse disponibles*

Le nombre d'études portant sur la méthylation de l'ADN a beaucoup augmenté depuis quelques années, grâce aux progrès technologiques. Ces études engendrent la production de données à haut débit, permettant aux chercheurs de comprendre les profils de méthylation ainsi que leur rôles régulateurs dans les processus biologiques menant aux maladies. Plusieurs logiciels permettent d'analyser ces données, mais il est encore difficile pour les praticiens de choisir la méthode la plus efficace. Le but de cette présentation est de comparer les outils statistiques disponibles au travers de données réelles et simulées, afin d'établir leur efficacité, leur validité et leur utilité.

---

**THIERRY CHEKOUO TEKOUANG**, UT MD Anderson Cancer Center

*Un modèle prédictif bayésien pour la génétique d'imagerie avec application à la schizophrénie*

Dans cet exposé, j'introduis un modèle bayésien intégratif de prédiction du risque qui permet de distinguer les patients schizophrènes des contrôles sains, basé sur un ensemble clairsemé de ROI (régions d'intérêt) et de SNP (polymorphisme d'un seul nucléotide) discriminantes. Une inférence sur un réseau régulateur entre les ROI et les SNP est utilisée dans une modélisation unifiée pour informer la sélection de ROI et de SNP discriminantes. En se servant d'études de simulations, nous évaluons la performance de notre méthode, et nous l'appliquons à des données de schizophrénie. Nous confirmons que certains bio-marqueurs impliqués dans le réseau ROI-SNP sont potentiellement discriminants.