



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Soutenance de thèse

Madame Sina RÜEGER

Titulaire d'un « Master » de l'Université des sciences appliquées de Zurich (ZHAW)

Soutiendra en vue de l'obtention du grade de
Doctorat ès sciences de la vie (PhD)
de l'Université de Lausanne

sa thèse intitulée :

**Integrative statistical analysis
of -omics and GWAS data**

Directeur de thèse :

Monsieur le Professeur Zoltan KUTALIK

Cette soutenance aura lieu le

Vendredi 21 septembre 2018 à 16h00

A la salle 2914 du Biophore, Quartier UNIL-Sorge, 1015 Lausanne

L'entrée est publique

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Résumé compréhensible

Titre de la thèse: Analyse statistique intégrative des données -omiques et GWAS

Nom doctorante: Sina Rüeger

Département: Département universitaire de médecine et santé communautaires

Approfondir nos connaissances concernant les mécanismes biologiques humains est essentiel pour aider la médecine à progresser, et permettre le diagnostic précoce de certaines maladies, le développement de médicaments, l'utilisation de stratégies de santé publique adaptées ou encore la mise en place de la médecine de précision. Pour cela, il est possible de mesurer les différentes composantes d'un mécanisme biologique et de faire le lien entre celles-ci pour essayer de construire un modèle sous-jacent qui aiderait à expliquer les causes de la maladie.

Dans le cas des études d'association pangénomiques (GWAS), des données génomiques (génotypage de variants génétiques) ainsi que le statut d'une maladie sont collectés pour un ensemble d'individus (cohorte). Ces données sont ensuite analysées, à l'aide d'outils statistiques, pour étudier l'association de chaque marqueur génétique avec le statut de la maladie au sein de la cohorte.

Cependant, ce type d'analyse ne permet pas d'identifier avec certitude les variants ayant un effet direct sur la maladie. Ces résultats de GWAS doivent être analysés plus en détails, en utilisant des méthodes supplémentaires, ou en les combinant avec d'autres types de données -omiques (notamment transcriptomique ou métabolomique) pour mieux comprendre les causes biologiques de la maladie.

Par exemple, une méthode statistique appelée randomisation mendélienne permet d'utiliser les résultats de GWAS de deux maladies, telles que les maladies cardiovasculaires et l'obésité, pour examiner l'effet de causalité de l'obésité sur les maladies cardiovasculaires, ou vice versa.

Ces analyses additionnelles nécessitent souvent des résultats de GWAS pour un large nombre de variants génétiques, et lorsque certains marqueurs n'ont pas été génotypés dans la cohorte, ces données ne sont pas disponibles.

Pendant mon doctorat, j'ai travaillé sur l'implémentation d'une méthode appelée imputation à partir de statistiques synthétiques qui vise précisément à résoudre ce problème. L'imputation à partir de statistiques synthétiques utilise les résultats de GWAS des marqueurs analysés (et non les données génétiques utilisées lors de l'analyse, qui ne sont généralement pas publiquement disponibles) et des données de séquençage publiques, « panel de référence », pour inférer les résultats de GWAS des variants manquants.

J'ai pu montrer que la précision de l'imputation varie en fonction des caractéristiques des variants génétiques (par exemple, faible précision pour les mutations rares) mais aussi en fonction de la taille du panel de référence utilisé (faible précision lorsque peu de séquences sont disponibles). J'ai également pu montrer que l'imputation à partir de statistiques synthétiques peut concurrencer les techniques d'imputation qui sont basées sur les données génomiques au niveau individuel pour certains sous-groupes de variants génétiques (notamment les variants communs).

L'imputation à partir de statistiques synthétiques peut faciliter l'utilisation de méthodes d'analyses additionnelles et ainsi aider les scientifiques à mieux comprendre les mécanismes et les causes biologiques de certaines maladies.