

M2 internship - Leveraging and modeling genetic pleiotropy in causal inference to understand the human genetic architecture

Keywords

statistical genetics; pleiotropy; post-genomic data; causal inference; Mendelian randomization; Gaussian mixture models

Description

A major transformation in genetics has occurred with the results of Genome-wide associations studies (GWASs). GWASs, which consist in estimating the effect of genome-wide genetic variants on a studied trait, have yielded to the identification of countless genetic variants significantly associated with many complex traits and diseases without being able to pinpoint a causal mechanism in the vast majority of cases. Therefore, many applications and method development have successfully reused the results of GWASs principally to study relationships between traits. One booming field using GWASs summary statistics data is causal inference between traits in the form of Mendelian randomization. The principle of Mendelian randomization is very simple and analogous to randomized control trials where the effects of variant alleles (instead of drug/placebo) are modeled through regression to estimate and test the causal effect of an exposure trait on an outcome trait. Although extremely appealing, Mendelian randomization relies on a strong assumption: the absence of horizontal pleiotropy occurring when a variant has independent effects on both the exposure and the outcome. Pleiotropy tended to be neglected in Mendelian randomization applications. In a stepping-stone paper published in *Nature Genetics* in 2018, we have shown that horizontal pleiotropy cannot be neglected and occurs in almost 50% of causal relationships, biasing causal estimates and inflating the false discovery rate of causal relationships. This internship will be dedicated to building a comprehensive method to perform Mendelian randomization and infer causality. In addition, we propose to reroute existing methods based on Mixture of Gaussians to infer causality while concomitantly quantifying pleiotropy at the level of the variants which we do believe to be primordial for our understanding of human genetics.

Importantly, the full code of the produced methodology and the results on pleiotropy will be made publicly available and highlighted in scientific publications.

References related to the internship

Verbanck, Marie, Chia-Yen Chen, Benjamin Neale, and Ron Do. 2018. "Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization Between Complex Traits and Diseases." *Nature Genetics*, April, 1. doi:[10.1038/s41588-018-0099-7](https://doi.org/10.1038/s41588-018-0099-7).

Darrous, Liza, Ninon Mounier, and Zoltán Kutalik. 2020. "Simultaneous Estimation of Bi-Directional Causal Effects and Heritable Confounding from GWAS Summary Statistics." *medRxiv*, January, 2020.01.27.20018929. doi:[10.1101/2020.01.27.20018929](https://doi.org/10.1101/2020.01.27.20018929).

Morrison, Jean, Nicholas Knoblauch, Joseph H. Marcus, Matthew Stephens, and Xin He. 2020. "Mendelian Randomization Accounting for Correlated and Uncorrelated Pleiotropic

Effects Using Genome-Wide Summary Statistics.” *Nature Genetics*, May, 1–8.
doi:[10.1038/s41588-020-0631-4](https://doi.org/10.1038/s41588-020-0631-4).

Successful candidate

The successful candidate:

- will have a master of data science linked to statistics or artificial intelligence, candidates with more theoretical background however showing strong interest in life science applications are also welcome;
- will be enthusiastic about transdisciplinary research and open science at the interface between data science and genetics;
- will show a clear interest to use applied science methodology to benefit biological understanding;
- will have good programming skills, preferentially R and/or Python;
- can have a background in biology or genetics;
- should be open-minded and willing to work as a team with other lab members;
- is expected to pursue PhD training as funding has already been secured through the PleioMap project (Funding: Agence Nationale de la Recherche, PI: Dr Marie Verbanck);
- will speak decent English since we are closely collaborating with Mount Sinai Hospital in New York City, USA.

NB: two master 2 internships will start concomitantly in the lab, although the two internships have very different aims and methodological angles, they have a similar background and are part of the PleioMap project, therefore both interns are expected to collaborate ([other internship offer](#)).

Scientific environment

Starting date: February 2022

The 6-months Master 2 internship will be supervised by Dr Marie Verbanck who is Assistant Professor (Maître de Conférences) at the Faculté de Pharmacie at Université de Paris within the UR 7537 BioSTM unit (Biostatistique, Traitement et Modélisation des données biologiques). BioSTM's mission is to develop cutting-edge statistical methodologies to answer real-life biological problems with an emphasis on reproducible science and open research.

How to apply

To apply, please send a concise email describing your research interests and experience as well as an up-to-date CV to Marie Verbanck (marie.verbanck@u-paris.fr). Name and contact for references will be appreciated.