

M2 internship - Modeling and characterizing genetic variant pleiotropy using machine learning to understand the human genetic architecture

Keywords

statistical genetics; pleiotropy; complex traits and diseases; post-genomic data; unsupervised and semi-supervised learning; penalized methods; random forests

Description

Nowadays in human genetics, one particular concept seems to resurge: pleiotropy. Pleiotropy occurs when one genetic element (e.g. variant, gene) has independent effects on several traits. Although pleiotropy is extremely common and thought to play a central role in the genetic architecture of human complex traits and diseases, it is one of the least understood phenomena. We have shown that several biological mechanisms exist and induce different pleiotropy states at the level of the variants. Specifically, we have conceptualized 5 biological mechanisms 1) linkage disequilibrium; 2) causality between traits; 3) genetic correlation between traits; 4) high polygenicity of traits; 5) horizontal pleiotropy (true independent effects of a variant on two traits). This internship will be dedicated to building a comprehensive framework to disentangle all 5 states of pleiotropy and provide a genome-wide map of pleiotropy using machine learning. Specifically, we propose 1) to improve on a method that we have published in a proof-of-concept paper using unsupervised approaches based on penalized methods, random forests or deep learning; 2) to explore semi-supervised learning using a creative strategy to label data that we have developed. There is a growing utility for Human genetic variant databases, from the interpretation of genetic analyses to clinical interpretation. We strongly believe that a database describing the pleiotropic nature of variants will complement existing databases and serve the community.

Importantly, the full code of the produced methodology and the genome-wide map of 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).

Jordan*, Daniel M., Marie Verbanck*, and Ron Do. 2019. "HOPS: A Quantitative Score Reveals Pervasive Horizontal Pleiotropy in Human Genetic Variation Is Driven by Extreme Polygenicity of Human Traits and Diseases." *Genome Biology* 20 (1): 222. doi:[10.1186/s13059-019-1844-7](https://doi.org/10.1186/s13059-019-1844-7).

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.