GEN-omix: integration of genetic information into multi-omics models (3 years PhD project)

Location: UMR1078/UMR6205, Brest, France

Supervision: Ozvan Bocher (ozvan.bocher@univ-brest.fr) and Vincent Calvez (vincent.calvez@math.cnrs.fr)

Starting date: between 01/10/2026 and 01/12/2026

Advances in sequencing and molecular profiling (also called 'omics') now make it possible to characterise biological traits (transcriptomics, proteomics, metabolomics, etc.) and study their role in human complex traits. Although each omic level provides information at a specific molecular level, integrating these data allows us to better capture their links and gain a more systematic understanding of biological mechanisms. Statistical methods have been developed to achieve this integration, such as DIABLO[1] and MOFA[2]. These methods are based on dimension reduction tools, which identify 'factors' that capture the variability of multi-omics data. However, these models only integrate genetic information to a very limited extent. This is especially due to the nature of the data, where genotype information does not correspond to quantitative metrics as for omics profiles, and presents a very high dimensionality. Yet, genome-wide association studies (GWAS)[3] have highlighted the crucial role of genetic variations in shaping the risk of human diseases such as type 2 diabetes[4]. Symmetrically, the genetic regulation of omics profiles is being increasingly characterized through QTL ('Quantitative Trait Loci') mapping[5]. Following integration of GWAS and QTL information at single-omics levels appears to be crucial to resolve the biology behind genetic associations with complex traits[6]. **The main objective of this thesis project is to integrate genetic information into multi-omics models to improve their performance and stability.** To achieve this objective, we propose to derive the genetic co-regulation between molecular traits based on publicly available information on QTLs, and to use it to inform multi-omics methods. This will require modifying the mathematical models underlying these methods. Two first approaches are envisioned: either changing a priori probabilities in the Bayesian inference framework, or the regularization of the cost function by the integration of a penalty to favour the use of QTL information in the model. To evaluate the genetically informed developed methods, a simulation framework will be developed to model various multi-omics settings. Additionally, benefits will be assessed on real data from collaborations of the team, offering insights into biological mechanisms from real world applications. These modifications and their contribution to current methods will be evaluated through simulations and on real data available through UMR1078 collaborations. Improving multi-omics models will ultimately reduce the required sample sizes, a key issue in multi-omics studies, which remain very costly to date. To make the newly developed method widely accessible, they will be implemented into the Biomix framework which aims at simplifying and popularizing the use of multi-omics models[7].

The simulations developed during this PhD project will also be made available to the community as a unified framework to assess multi-omics methods is highly needed in this rapidly expanding field. In order to carry out this transdisciplinary project, the PhD student will benefit from the expertise of two research labs involved in the supervision of the project. **UMR1078** has a strong expertise in genetics, genetic epidemiology and omics analysis. The PhD project will be mainly supervised by Ozvan Bocher, junior professor in this research unit. Researchers from **UMR6205** (Vincent Calvez and François Ged) will provide expertise in Bayesian modelling and applied mathematics. During this PhD, the candidate will benefit from a co-supervision and a stimulated research environments with all the necessary infrastructure from both teams. They will have the opportunity to attend disciplinary seminars and to present their work to national and international conferences.

Candidate profile:

- Master's degree in a quantitative field (e.g. computational biology, data science, applied mathematics, etc.)
- Skills in genetics/genomics epiodemiology, applied mathematics/statistics (e.g. high-dimensional analysis, Bayesian methods, etc.)
- Computational skills: proficiency in Python or R
- Very good written and oral **English** skills
- Experience in data analysis, particularly multi-omics, would be a major asset

References:

1. Singh A, Shannon CP, Gautier B, et al. *Bioinformatics*. 2019;35(17):3055-62.
2. Argelaguet R, Velten B, Arnol D, et al. *Molecular Systems Biology*. 2018;14(6):e8124.
3. Abdellaoui A, Yengo L, Verweij KJH, et al. *The American Journal of Human Genetics*. 2023;110(2):179-94.
4. Suzuki K, Hatzikotoulas K, Southam L, et al. *Nature*. 2024;627(8003):347-57.
5. Aguet F, Alasoo K, Li YI, et al. *Nature Reviews Methods Primers*. 2023;3(1):4.
6. Bocher O, Arruda AL, Yoshiji S, et al. *Nature Metabolism*. 2026;8(2):506-20.
7. Iperi C, Fernández-Ochoa Á, Barturen G, et al. *BMC Bioinformatics*. 2025;26(1):8.