

ID03:

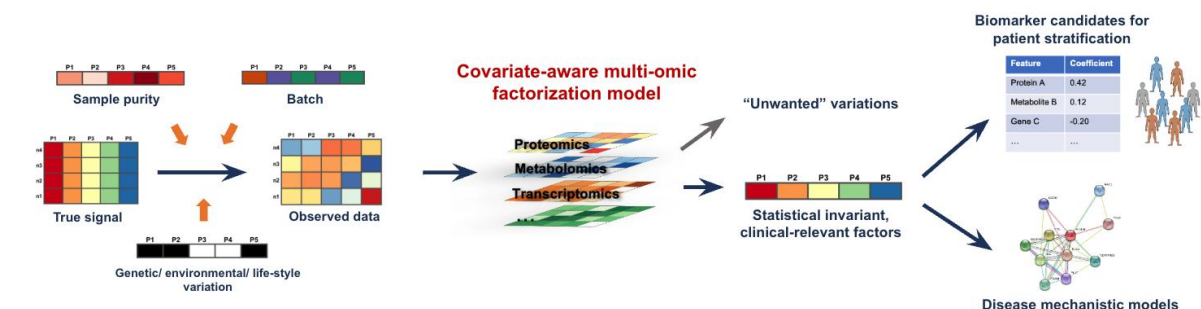
Agentic, covariate-aware multi-omics factor analysis for causal biomarker discovery

Data Science supervisor: Prof. Dr. Oliver Stegle

Life Science supervisor: Dr. Junyan Lu

Multi-omics profiling promises a comprehensive view of disease processes across molecular layers, with broad implications for personalised medicine. Despite this potential, the analysis of high-dimensional multi-omic data remains a major theoretical and practical challenge. Clinical multi-omic readouts are noisy, sparsely sampled, and shaped by complex covariate structures, including genetic background, technical variation (e.g. batch effects), temporal and spatial sampling, and experimental conditions such as chemical or genetic perturbations. Together, these factors can preclude the identification of causal and mechanistic signals. Multi-omics factor analysis, pioneered by the partnering PI groups, is among the most promising strategies for distilling biomedical insights from such data. However, existing methods cannot adequately account for the complex covariate structures of clinical cohorts, and they are difficult to specify, adapt, and deploy without substantial methodological expertise, which severely limits their uptake by clinical scientists.

To address these challenges, we propose to develop a flexible, programmatic, covariate-aware multi-omics factorisation framework tailored to heterogeneous clinical cohorts with complex experimental designs. The framework will advance the state of the art through automatic model definition, structured handling of patient and experimental covariates, and strategies that exploit instrumental variables and statistical invariance to enable causal representation learning. In parallel, the framework will be paired with agentic interfaces that make model specification, configuration, and interpretation accessible to clinical scientists without prior expertise in probabilistic modelling. We will apply the resulting solutions to a comprehensive multi-omics cohort of lung cancer patients to identify biomarkers predicting tumor recurrence, the primary cause of treatment failure and cancer-related death.



[Prof. Dr. Oliver Stegle](#)



[Dr. Junyan Lu](#)

Special requirements for applicants:

- Master's degree or equivalent degree in bioinformatics, informatics, computational biology, or related subjects.
- Strong programming skills in at least one programming language, preferably in Python or R.
- Track record in probabilistic modeling and/or machine learning, preferably experiences with probabilistic programming frameworks.
- Good knowledge and understanding of the concepts of cancer biology, personalized medicine, and biomarker discovery.
- Experience with the analysis of high-throughput omics datasets (e.g. proteomics, transcriptomics, genomics and etc.) is a bonus.