Project Proposals for Doctoral Researcher Positions 2025

ID06: Enhancing Biomarker Discovery through Covariate-Aware Multi-Omics Factorization Models (Oliver Stegle, Junyan Lu)

Heidelberg, DKFZ/EMBL, Computational Genomics and Systems Genetics

Multi-omics studies hold promise for enhancing our understanding of biological processes across multiple molecular layers, with significant implications for personalized medicine. However, analyzing high-dimensional multi-omic data remains challenging due to noise, incomplete sampling, and confounding sources of variation, limiting the ability to derive causal insights.

Matrix factorization methods, like multi-omics factors analysis pioneered by partnering PI groups, are promising for extracting biomedical insights from such data. Yet, current methods often overlook critical patient sample covariates such as genetic backgrounds, technical variations (e.g., batch effects), temporal/spatial structures, and experimental conditions (e.g., perturbations), hindering their applicability in clinical settings. Moreover, there's a lack of adaptable multi-omics models tailored to specific datasets, further impeding clinical adoption.

To address these gaps, this project proposes developing a flexible, programmable covariate-aware multi-omics factorization model. This model will be designed to analyze heterogeneous clinical samples with complex covariate structures, advancing multi-omics analysis by automating model definition and incorporating strategies to enhance causal representation learning using instrumental variables.

The project will apply these advancements to a comprehensive multi-omics cohort of lung cancer patients. The primary objective is to identify biomarkers predictive of tumor recurrence, a critical factor in treatment efficacy and patient outcomes in cancer care. This research aims to propel multi-omics analysis forward, facilitating its integration into clinical practice for improved patient stratification and personalized treatment strategies.

Requirements:

- 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.

Oliver Stegle: https://www.dkfz.de/bioinformatik-der-genomik-und-systemgenetik

Junyan Lu: https://lu-group-ukhd.github.io/dr-lu/