

Thesis project: Large scale learning of epigenetic regulation from paired scRNA and scATAC data

Importance of epigenetic regulation: Epigenetic regulation is central to basic cellular processes, such as transcription and post-transcriptional regulation, and thus important in human development and disease. In addition to genetically inherited mutations, epigenetic variation contributes to onset and progress of diseases. Our understanding of epigenetic variation in disease rests on analysis of genome-wide data that allows to contrast several cell-type and disease conditions and learn computational models that allow their interpretation.

Experiment and data; what do we want to investigate? We routinely work on methods that use epigenetics and expression data to learn about epigenetic regulation of genes. We have access to a large scale single cell dataset and want to improve our way to understand how epigenetic variation contributes to disease. The dataset contains several disease models.

Approach: Using supervised Machine learning the task is to establish a relationship between epigenetic activity in genomic regions and expression of genes. As the data is from single cell data, it is able to capture purity of different cell types in disease but also it is very noisy and therefore challenging in its application.

What does the student learn: The project would allow the applicant to learn how to estimate machine learning models for large datasets, infer cell type gene-regulatory regions and assess how they connect to disease.

Requirements: Existing knowledge of R or python (preferably), machine learning and statistical modelling or keen interest in learning these.