Thesis project: Machine Learning for detection of patterns of alternative polyadenylation in time series data

Importance of miRNA and miRNA regulation: microRNAs (miRNA) are small non-coding RNA molecules that can extensively post transcriptionally regulate gene expression. They bind to their target messenger RNA (mRNA) transcripts, thereby repressing their expression in the cell.

Importance of alternative polyadenylation: However, mRNA transcripts undergo extensive modifications as well. Cellular stress is known to further induce these modifications through splicing and alternative polyadenylation (APA), adding diversity to the transcriptome. This leads to the gain or loss of miRNA binding sites depending on modifications such as lengthening or shortening of the transcript. Thus lengthening of mRNA results in additional repression of the transcript while shortening results in an escape from repression.

Experiment and data; what do we want to investigate? We performed small and total RNA sequencing from four cell types of the murine heart at six timepoints following induced acute myocardial infarction. From our initial investigations we found transcripts exhibit PA modifications overtime and miRNAs show interesting expression dynamics in each cell type. In this part of the project, we would like to study the dynamics of miRNA regulation and transcript polyadenylation response upon stress.

Approach: First, we predict miRNA binding sites on polyadenylated transcripts over time. Using these predicted miRNA binding sites on PA transcript isoforms, we build a machine learning model so as to predict miRNA expression using the transcript binding sites and transcript isoform expression. We then compare predictability of transcripts undergoing modifications and containing (or devoid) of target sites to transcripts with no modifications. Comparing the miRNA regulatory patterns, we would like to investigate if polyadenylation interferes with miRNA mediated transcript regulation, upon stress. Moreover, comparison of miRNA regulation between cell types, would also provide information on target pathways for their role in adaptability of the heart cell types upon stress.

What does the student learn: The project would allow the applicant to model complex gene-regulation pathways, infer cell type specific miRNA mediated effects on the biological system and presentation of the results.

Requirements: Existing knowledge of R or python (preferably), machine learning and statistical modelling or keen interest in learning these. Knowledge of miRNA and transcription biology is a plus.