Prof. Dr. Christian L. Müller
Ludwig-Maximilians-Universität München
Institut für Statistik
Ludwigstr. 33
80539 München
christian.mueller@stat.uni-muenchen.de

M.Sc. Thesis Proposal: *Methods for the detection of gene expression induction in response to chemical stress*

Objective: The main objective of this M.Sc. thesis proposal is to develop a statistical pipeline that detects chemical stress that significantly increase (or decrease) the expression of bacterial pathogen genes. The main objectives include: i) explore different approaches for data pre-processing for lux curve normalisation, ii) compare different methods for detecting significant changes in gene expression such as non-parametric cutoffs, analytical p-values and empirical p-values, and iii) *post-hoc* analysis of determined "hits" with network inference.

To accomplish this, we will make use of the data generated by the StressRegNet [1] consortium. The student can access pre-made pipelines to start their work, which explores hit determination by summarizing gene expression time-series as area under curves (AUC) and uses analytical p-values. The student will explore the analysis from the perspective of various functional data analysis techniques as well.

Once "hits" are determined, the student will analyze the similarity between compounds based on their measured effects on gene expression via various clustering and network inference algorithms. At the same time, detection of genetic regulatory circuits will be carried out with similar methods.

Plan and deliverables: A successful completion of the M.Sc. thesis requires the following computational and scientific advances. Firstly, the thesis should deliver a workflow that allows for the analysis of paired growth and lux curves. The results should be visualized in a dashboard. Secondly, the student will use the output of this workflow in order to detect chemical species with reproducible effects on gene expression, as well as the regulatory relationships between genes.

A possible outcome of this analysis would be the determination of empirical p-values as the best method for "hit" determination when considering growth-normalized AUC values. Based on these normalized expression values, we use GLASSO to determine

that all non-steroidal anti-inflammatory drugs and polyphenols activate efflux pumps. At the same time, via matrix inversion, we find that *marA* regularly regulates pump expression in response to chemical stress.

References

[1] Bayresq (n.d.). StressRegNet: A chemical-genomics approach to decipher stress response and virulence pathways in infection. *Bayresq.net*. https://bayresq.net/en/projekte-stressregnet-en/.