

## Phosphatase Activity Enrichment Analysis (PAEA)

Linda Kadamala Samuel

Master thesis in Bioinformatics and Computational Biology

The altered activity of kinase and phosphatase enzymes was shown to be characteristic of many diseases. Tyrosine kinases for example have been found to be overexpressed in different cancer types such as chronic myelogenous leukemia (CML), breast cancer and colorectal cancer. As a result, the study of phospho-signalling pathways continues to be of great interest to develop target-specific drugs. Hallal et al. (2020) at the University of Bern recently developed the Kinase Activity Enrichment Analysis (KAEA) pipeline to infer the differential activity of kinases from mass spectrometry-based phosphoproteomics data.

The aim of this project is to investigate the extension of the pipeline to include the enrichment analysis of phosphatases, which catalyse the dephosphorylation of proteins and thus have the opposite function of kinases. The enrichment analysis was performed with the SetRank algorithm, which is a highly specific method to identify significant gene sets. The phosphatase database for SetRank was created with experimentally verified data from the Human Dephosphorylation Database DEPOD. Additionally, the dataset was complemented with in-silico phosphatase-substrate predictions. Artificial neural network models for the four largest phosphatase superfamilies were trained with the data from DEPOD. The models were applied on three phospho-enriched human myeloid cell lines.

A considerable increase in data coverage was observed by complementing the DEPOD dataset with the neural network predictions. Altogether this thesis provides a useful resource for a better understanding of dephosphorylation reactions by enhancing experimental data with computational predictions.

PD Dr. Manfred Heller, Dr. Anne-Christine Uldry