Bioinformatics Analysis, Management and Organization of Biological Data related to Post-Translational Regulation.

Post-translational regulation is an important, fast and energy efficient level of gene regulation that has attracted the focus of many high-throughput technologies in the last 20 years. Post-translational modifications of amino acids and especially protein phosphorylation play a pivotal role at this level of cellular regulation. Accordingly, this thesis focused on publicly available and abundant high-throughput protein phosphorylation and methylation data, in order to develop computational tools and bioinformatics methods and pipelines, with the aim to analyze them and transform raw data into biological knowledge, about the properties of the eukaryotic phosphoproteome. During this thesis, phosphoproteomic and methylproteomic data were mined from the literature. An annotation tool and a database were developed in order to facilitate the mining and storage of these complex data, that were integrated with many other omic and evolutionary data. Statistical analyses of the gathered and filtered data allowed for a reliable estimate of the total number of phosphoproteins and phosphorylation sites in model eukaryotes. Furthermore, a focused and in-depth study of the yeast phosphoproteome revealed its pivotal role in the central metabolism and further identified key metabolic processes of biotechnological importance that may be manipulated in the future, with precision, by mutating key phosphorylation sites. Finally, neural networks were developed to predict phosphorylation and methylation sites and further predict potential meth-phos switches and/or clusters. The tools and analyses that were developed during this thesis may function as the first step towards more advanced tools and methods that will integrate many other post-translational modifications in the future.