

Omics data integration

Title: Exploring integration of metabolomics data into ProteomicsDB

Type: [MA]

Category: [Data Science | Database]

Programming language: [any + SQL]

Language: [English]

Prior experience: [experience with SQL and a programming language (e.g. Python) is required, biological knowledge is a plus but not necessary]

Complexity/Risk: [medium | high]

Contact person: Ludwig Lautenbacher and Mathias Wilhelm

Brief background description: Metabolomics, the large-scale study of metabolites, is quickly gaining in popularity as yet another layer of biochemical information. These small molecules can give us information about the current state of a cell and its interactions with proteins are fundamental for biological pathways. As these small molecules do not have the same regular chain structure as RNA and proteins, they have traditionally been much harder to analyze and their coverage has therefore been rather low until recently. Recently, many efforts have been made to integrate proteomics and metabolomics, though the integration has been rather superficial and inconclusive regarding the question of whether we gain more information by combining the two.

Literature:

- Wishart, David S., et al. "HMDB 3.0—the human metabolome database in 2013." *Nucleic acids research* 41.D1 (2012): D801-D807.
- Patel, Vishal R., et al. "CircadiOmics: integrating circadian genomics, transcriptomics, proteomics and metabolomics." *Nature methods* 9.8 (2012): 772.
- Ma, Yanlei, et al. "An integrated proteomics and metabolomics approach for defining oncofetal biomarkers in the colorectal cancer." *Annals of surgery* 255.4 (2012): 720-730.

Brief description of the project: In ProteomicsDB, we have large volumes of tissue-specific information with regard to transcriptomics and proteomics. However, currently, no metabolomics data is integrated yet. In this project, you will explore what the benefits would be of including metabolomics data in ProteomicsDB and what kind of analysis tools would be needed to make this information useful. Also, you will investigate the status-quo of methods for integrating proteomics and metabolomics and evaluate which methods would be most suitable for

ProteomicsDB. You will also have the opportunity to expand the existing multi-omics data model of ProteomicsDB to support quantitative metabolomic data.

Expected result: A proof of concept for including large-scale metabolomics data in ProteomicsDB, which could lead to ProteomicsDB becoming a leading resource for the integration of these two modes.