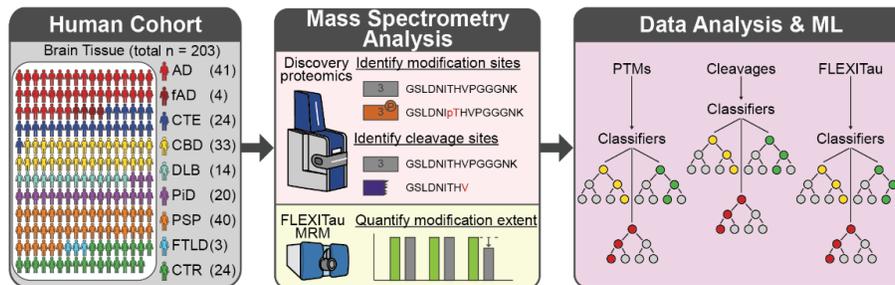


Project description:

This bachelor project aims to expand a data analysis platform for protein data analysis with a focus on statistics and machine learning capabilities. The project's underlying idea is to provide a one-stop-shop for the handling and the analysis of protein level data to aid biomedical researchers in their tasks to find changes in proteins associated with disease. The data source is a million-dollar fine-tuned scale, called mass spectrometer, which is used to analyze proteins in biological samples in healthy and diseased cells. By applying suitable statistical and machine learning methods in the analysis and appropriately visualizing the complex results, differences that correlate with disease can be identified and utilized for diagnosis or treatment of disease.



Workflow from biological samples through mass spectrometry proteomic analysis to data analysis and machine learning. This workflow enables the identification of molecular changes between healthy (CTR) patients and patients suffering from diseases, such as Alzheimer's disease (AD), Chronic Traumatic Encephalopathy (CTE) and others.

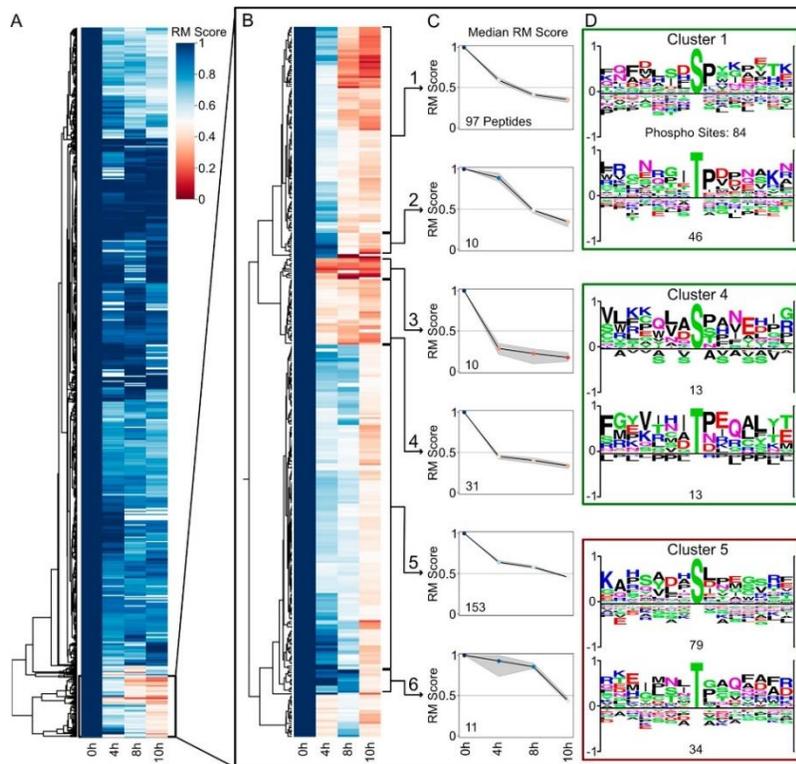
The individual analysis components are driven by the study design and biomedical questions for available data. The Steen&Steen laboratories at Boston Children's Hospital (BCH) are dedicated to understanding disease through changes in proteins. They generate large-scale proteomics datasets to answer brain questions surrounding the immune system, changes in proteins due to SARS-CoV-2, and neurodegenerative diseases, such as Alzheimer's disease and Parkinson's disease. Their expertise in mass spectrometry instrumentation and biochemistry enables the development of new laboratory-based techniques and the upscaling of studies. As new types of proteomics data and study setups become available, a compact and flexible analysis platform supporting robust statistical and machine learning analyses is essential. HPI and BCH have collaborated to create the initial framework and will continue to expand the software further.

The requirements listed below are to be implemented and added to the existing framework:

- Support of multi-level input data (proteins, peptides and post-translational modification);
- Various different statistical tests and machine learning approaches for analysis;
- Software implementation using agile development methodology;
- Optimized visualization of simple and complex results.

What you should bring to the table for the project:

- Knowledge in programming in python or strong software development knowledge and ability to transfer to python;
- Willingness to interact directly with the project partner;
- Interest in the biological application field (no prior knowledge required);
- First prior contact with statistics and/or machine learning is desirable but not required.



Example of analysis and visualization: (A) Hierarchical clustering analysis of peptides and their changes during the cell cycle. (B) Focus on a subgroup with increasing changes over time (C) Grouped visualization of the changes of subclusters over time and (D) the identification of sequence similarities (motifs) of peptides similarly changed over time in clusters.

Project partner:

Boston Children's Hospital (BCH) is the leading pediatric hospital in the United States. The Steen&Steen laboratories at the hospital are interested in understanding molecular, such as proteins, disease mechanisms and utilize this knowledge to advance diagnosis and treatment of diseases ranging from allergies to neurodegeneration. They employ novel mass spectrometry technology that require novel and performant implementation of analysis methods and novel visualization.

Supervisors from the Data Analytics and Computational Statistics chair:

Dr. Christoph Schlaffner – Christoph.Schlaffner@hpi.de

Prof. Dr. Bernhard Renard – Bernhard.Renard@hpi.de

Please don't hesitate to come and see us in F-E4 and F-E5 on campus II, to send us an email or set up a zoom meeting.