

**Veni, Vidi, Visualization:
Improving analysis communication for a-million-dollar machine**

Project description:

This bachelor project is an expansion to our data analysis platform PROTzilla, developed at our chair, for proteins and their changes and focuses on the complex and interactive visualization of methods and analysis results. The tool helps researchers identify changes of proteins between healthy and disease states utilizing data from a 1-2 million-dollar, fine-tuned scale, called mass spectrometer. PROTzilla is a one-stop-shop for handling the analysis of these protein-based data with suitable statistical and machine learning methods and communicating the results is primarily done through data visualization. By providing an interactive approach to the visualization of complex data, researchers can streamline their analysis, identify proteins involved in disease quicker, and can easily communicate their results to their peers at conferences and in publications helping find treatment or cures for various diseases.

The individual visualization types are driven by study design, biomedical questions and the analyses performed. The Functional Proteomics Group at The Institute of Cancer Research is dedicated to understanding cancer through changes in proteins. They generate large-scale proteomics datasets to answer questions surrounding interactions of proteins with one another and identifying new, never-before-seen proteins that only occur in cancer that can lead to the development of new cancer treatments. Their expertise in mass spectrometry, biochemistry and bioinformatics enables the advancement of laboratory-based techniques, the upscaling of studies, and the development of new computational analysis approaches. With new types of proteomics data, study setups, and analysis methods being developed, the flexibility and extensibility of PROTzilla to provide robust statistical and machine learning analyses and a strong set of adaptable visualization tools is essential. ICR has been a beta-tester of PROTzilla in the past and collaborates with HPI to expand the software.



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

- Support of multi-level data visualization (proteins, peptides and post-translational protein modification that are interconnected);
- Develop concepts for improved user interaction/experience;
- Software implementation using agile development methodology;
- Optimized visualization for data exploration and publishing.

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;
- Strong interest in data visualization or visual analytics (no prior knowledge required);
- Interest in the biological application field (no prior knowledge required);
- First prior contact with data visualization and/or statistics or machine learning is desirable but not required.

Project partner:

The Institute of Cancer Research (ICR) as an institute has been involved in the most important discoveries in the history of cancer research. The Functional Proteomics research group, led by Prof. Jyoti Choudhary, is interested in understanding molecular mechanisms, specifically of proteins, of different cancers and utilizes this knowledge to advance diagnosis and treatment of all kinds of cancers. They employ novel mass spectrometry technology that require novel and performant implementation of analysis methods and novel visualization to generate new knowledge and communicate their findings to colleagues and the world in general.

Supervisors from the Data Analytics and Computational Statistics chair:

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

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

Simon Witzke – Simon.Witzke@hpi.de

Please do not hesitate to come and see us in K-E.15, K-E.14 and K-E.9/10 on campus I, to send us an email or set up a zoom meeting.

Join us!

