
Subject Section

ProteomicsBrowser: MS/Proteomics Data Visualization and Investigation

Gang Peng^{1,2}, Rashaun Wilson³, Yishuo Tang², TuKiet T. Lam^{4,5}, Angus C. Nairn³, Kenneth Williams⁴ and Hongyu Zhao^{1,*}

¹Department of Biostatistics, School of Public Health, Yale University, New Haven, CT 06510, USA, ²Department of Genetics, School of Medicine, Yale University, New Haven, CT 06510, USA, ³Department of Psychiatry, School of Medicine, Yale University, New Haven, CT 06510, USA, ⁴Department of Molecular Biophysics and Biochemistry, School of Medicine, Yale University, New Haven, CT 06510, USA, ⁵MS & Proteomics Resource of Keck Biotechnology Resource Laboratory at Yale, New Haven, CT 06510, USA.

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on 10 September 2018; revised on 26 October 2018; accepted on 20 November 2018

Abstract

Summary: Large-scale, quantitative proteomics data are being generated at ever increasing rates by high-throughput, mass spectrometry technologies. However, due to the complexity of these large datasets as well as the increasing numbers of post-translational modifications (PTMs) that are being identified, developing effective methods for proteomic visualization has been challenging. ProteomicsBrowser was designed to meet this need for comprehensive data visualization. Using peptide information files exported from mass spectrometry search engines or quantitative tools as input, the peptide sequences are aligned to an internal protein database such as UniProtKB. Each identified peptide ion including those with PTMs are then visualized along the parent protein in the Browser. A unique property of ProteomicsBrowser is the ability to combine overlapping peptides in different ways to focus analysis of sequence coverage, charge state, or PTMs. ProteomicsBrowser includes other useful functions, such as a data filtering tool and basic statistical analyses to qualify quantitative data.

Availability and implementation: ProteomicsBrowser is implemented in Java8 and is available at <https://medicine.yale.edu/keck/nida/proteomicsbrowser.aspx> and <https://github.com/peng-gang/ProteomicsBrowser>.

Contact: hongyu.zhao@yale.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Rapid advances in high-throughput MS instrumentation have resulted in an exponential growth of large-scale, quantitative proteomics studies during the past decade (Efstathiou, et al., 2017). The demands for high-performance data processing and proteomics analysis software are greater today than ever before, as many more research groups have the capacity and capability to generate complex proteomics data sets on a daily basis. As noted by Gatto et al. (Gatto, et al., 2015), visualization plays an essential

role in high-throughput biology. It is very important that the software used to interpret MS/MS proteomics data has the ability to display key findings both quantitatively *and* visually. While there have been enormous advances in LC/MS/MS instrumentation and in the automation of data acquisition, it has been observed (Avtonomov, et al., 2016) (Efstathiou, et al., 2017) that software solutions for “bottom-up” proteomics data processing, analysis, and visualization have lagged behind. In this regard, overlapping peptides resulting from partial proteolytic cleavage, multiple charge states, and sub-stoichiometric post-translational modifications

(PTMs) can easily result in >20 peptides covering a PTM site of interest. It may be helpful to combine overlapping peptides based on their sequence, charge state, or PTMs. Since a review of ~30 programs (Supplementary Table 1) for the analysis and visualization of MS/proteomics data sets did not identify any with these capabilities, we designed ProteomicsBrowser to meet these needs.

2 Methods

ProteomicsBrowser is coded in Java8 (<https://www.java.com/>) that can run on multiple platforms including Windows, macOS, and Linux. ProteomicsBrowser takes comma-separated values (CSV) files with peptide information exported from a mass spectrometry search engine or peptide quantitative output from third party vendors as input. ProteomicsBrowser searches the internal database to locate the position of each peptide ion in a protein. The internal database uses fasta format to store protein sequences for different organisms, which can be updated by downloading protein information from UniProt (<http://www.uniprot.org>).

3 Results

As described in the User Guide in Supplementary Information and Supplementary Figure 1, after importing LC/MS/MS data from a study that typically contains multiple “control” versus “experimental” samples, ProteomicsBrowser organizes the project into two sections: Data and Browser. The Data section allows users to view the peptide and protein data in a table and to perform some statistical analyses. The key function of ProteomicsBrowser is its Browser that was designed to help users visualize peptides in a protein of interest in any one of the samples in the project being analyzed. As shown in Fig. 1, the Browser depicts each peptide as a horizontal gray bar (box) above the corresponding sequence. The peptide ions are aligned along the parent protein with overlapping peptides being stacked vertically as shown in Fig. 1. The intensity of the gray color of each bar indicates the relative peptide abundance with color-coded vertical lines in the bars indicating the positions and type of PTMs that were identified in the selected protein. A particularly novel feature of the Browser is its ability to simplify the depiction of related peptides by combining peptide ions (Fig. 1) based on their sequence, charge, and/or specific PTMs. In addition, the unique “Quantify PTM” option can be used to rapidly determine the extent of post-translational modification at a selected residue in a protein of interest by combining all of the overlapping peptide ions that contain the PTM of interest into one group and all of the overlapping peptide ions that do not contain that PTM into a second group. The integrated areas of the peptides with the selected PTM are summed and depicted as a single peptide box. Similarly, the overlapping peptides without the PTM of interest are combined into a different peptide box. The extent of modification can be easily calculated from the ratio of the sum of the integrated peak areas of the combined PTM peptides to the sum of the combined PTM + non-PTM peptides. Another unique feature of ProteomicsBrowser is its ability to export a text file or figure that shows the amino acid frequencies around a PTM of interest similar to a BlockLogo tool (Olsen, et al., 2013). All of the results and Browser views can be exported to text files or figures to facilitate publication.

4 Discussion

ProteomicsBrowser has unique abilities to clearly and easily visualize the positions and relative abundances of all of the identified and quantified peptides and their PTMs from a selected protein. This simplified depiction based on sequence, charge, and/or PTM should greatly aid in the analysis

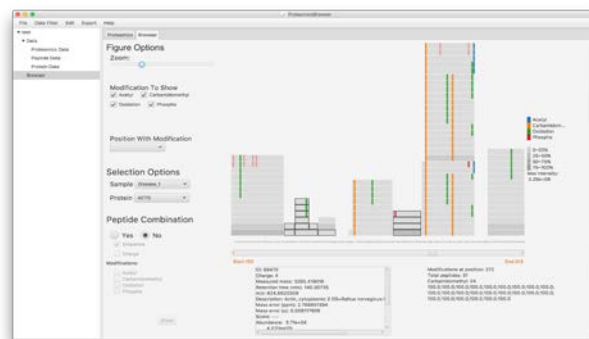


Fig. 1. GUI of ProteomicsBrowser. Peptide ions identified in the ACTG protein from the Disease-1 sample. The left sidebar allows the user to customize and control the analysis procedure, including using the Selection Options to choose a particular sample, a particular protein in that sample, and changing the scale of visualization with the Zoom control. The center panel presents the overall visualization of the alignment of the identified and quantified peptides. As shown, with the Zoom control near the mid position, it is possible to visualize residues 150-313 as designated by the red font “Start 150” on the left and the “End 313” on the right.

and interpretation of complex LC/MS/MS data sets. In particular, ProteomicsBrowser will leverage recent advances in LC/MS/MS instrumentation by facilitating the optimal selection of proteotypic peptide ions for designing targeted proteomics analyses such as Parallel Reaction Monitoring (PRM) (Rauniyar, 2015), and in the determination of the overall extent of modification of individual PTMs. These capabilities are augmented by the inclusion in the ProteomicsBrowser of selected statistical analyses tools and a wide range of options for filtering the data. We will continue to improve the ProteomicsBrowser and to add new features based on user feedback and will post the latest version at <https://medicine.yale.edu/keck/nida/proteomicsbrowser.aspx>.

Acknowledgements and Copyright

We thank Drs. Melissa Monsey and Jane Taylor for use of their data. The ProteomicsBrowser is copyrighted by Yale University.

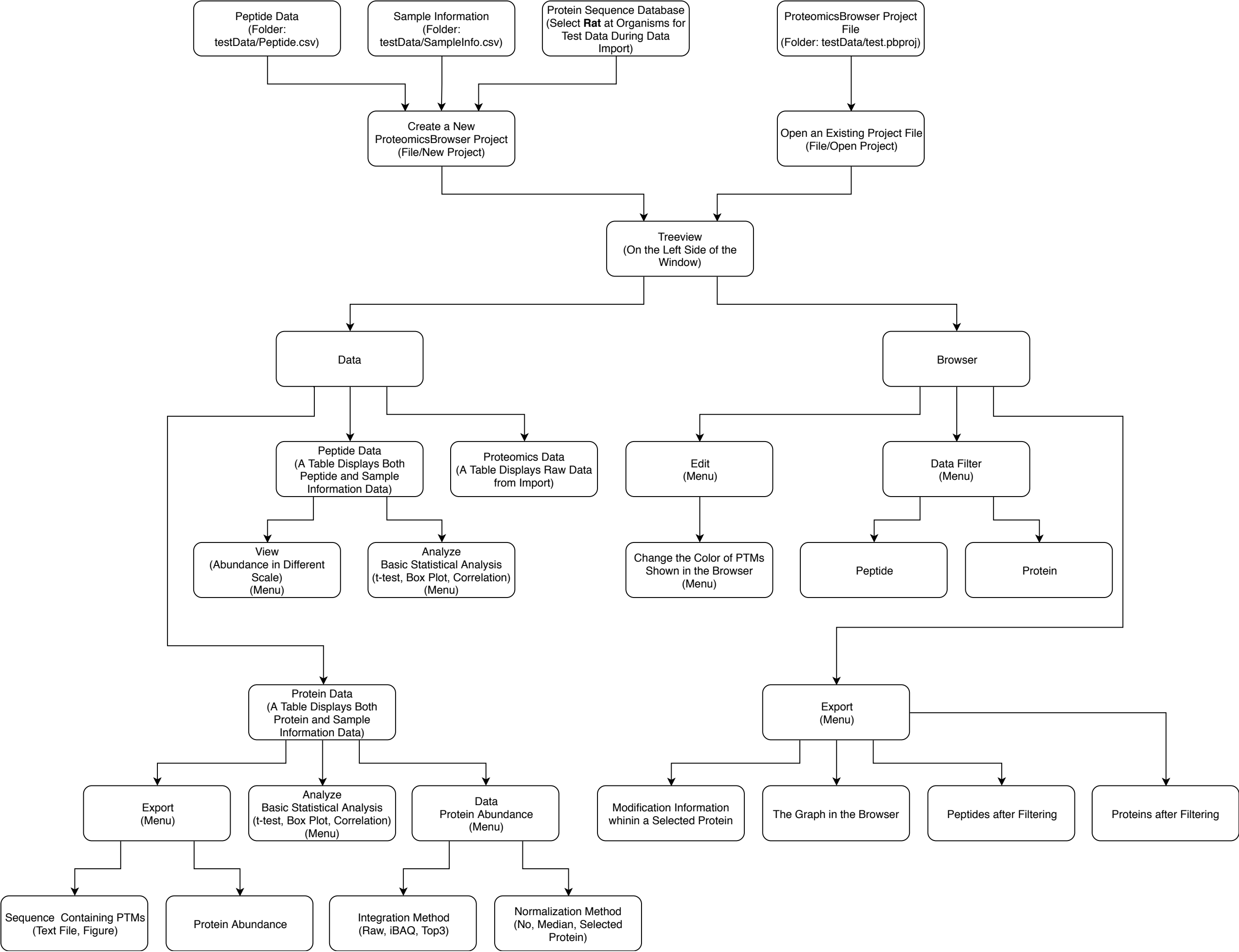
Funding

This work has been supported by NIH grant P30 DA018343. The LC MS/MS data utilized for testing were collected on an NIH SIG-supported mass spectrometer (1S10DOD018034).

Conflict of Interest: none declared.

References

- Avtonomov, D.M., Raskind, A. and Nesvizhskii, A.I. BatMass: a Java Software Platform for LC-MS Data Visualization in Proteomics and Metabolomics. *Journal of proteome research* 2016;15(8):2500-2509.
- Efstathiou, G., et al. ProteoSign: an end-user online differential proteomics statistical analysis platform. *Nucleic acids research* 2017;45(W1):W300-W306.
- Gatto, L., et al. Visualization of proteomics data using R and bioconductor. *Proteomics* 2015;15(8):1375-1389.
- Olsen, L.R., et al. BlockLogo: visualization of peptide and sequence motif conservation. *Journal of immunological methods* 2013;400-401:37-44.
- Rauniyar, N. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *International journal of molecular sciences* 2015;16(12):28566-28581.



Supplementary Table 1: References for Software Programs for the Analysis and Visualization of MS/Proteomics Data

- Aiche, S., *et al.* Workflows for automated downstream data analysis and visualization in large-scale computational mass spectrometry. *Proteomics* 2015;15(8):1443-1447.
- Avtonomov, D.M., Raskind, A. and Nesvizhskii, A.I. BatMass: a Java Software Platform for LC-MS Data Visualization in Proteomics and Metabolomics. *Journal of proteome research* 2016;15(8):2500-2509.
- Barsnes, H., Mikalsen, S.O. and Eidhammer, I. MassSorter: a tool for administrating and analyzing data from mass spectrometry experiments on proteins with known amino acid sequences. *BMC bioinformatics* 2006;7:42.
- Bielow, C., Mastrobuoni, G. and Kempa, S. Proteomics Quality Control: Quality Control Software for MaxQuant Results. *Journal of proteome research* 2016;15(3):777-787.
- Bryan, K., *et al.* HiQuant: Rapid Postquantification Analysis of Large-Scale MS-Generated Proteomics Data. *Journal of proteome research* 2016;15(6):2072-2079.
- Chen, T., *et al.* Web resources for mass spectrometry-based proteomics. *Genomics, proteomics & bioinformatics* 2015;13(1):36-39.
- Deutsch, E.W., *et al.* Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics. Clinical applications* 2015;9(7-8):745-754.
- Efstathiou, G., *et al.* ProteoSign: an end-user online differential proteomics statistical analysis platform. *Nucleic acids research* 2017;45(W1):W300-W306.
- Gatto, L., *et al.* Visualization of proteomics data using R and bioconductor. *Proteomics* 2015;15(8):1375-1389.
- Gehlenborg, N., *et al.* Prequips--an extensible software platform for integration, visualization and analysis of LC-MS/MS proteomics data. *Bioinformatics* 2009;25(5):682-683.
- Guner, H., *et al.* MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *Journal of the American Society for Mass Spectrometry* 2014;25(3):464-470.
- Haga, S.W. and Wu, H.F. Overview of software options for processing, analysis and interpretation of mass spectrometric proteomic data. *Journal of mass spectrometry : JMS* 2014;49(10):959-969.
- Hartler, J., *et al.* MASPECTRAS: a platform for management and analysis of proteomics LC-MS/MS data. *BMC bioinformatics* 2007;8:197.
- Jaitly, N., *et al.* Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC bioinformatics* 2009;10:87.
- Kroll, J.E., *et al.* A tool for integrating genetic and mass spectrometry-based peptide data: Proteogenomics Viewer: PV: A genome browser-like tool, which includes MS data visualization and peptide identification parameters. *Bioessays* 2017;39(7).
- Manguy, J., *et al.* Peptigram: A Web-Based Application for Peptidomics Data Visualization. *Journal of proteome research* 2017;16(2):712-719.
- Monroe, M.E., *et al.* MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Computational biology and chemistry* 2008;32(3):215-217.
- Mortensen, P., *et al.* MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *Journal of proteome research* 2010;9(1):393-403.
- Perez-Riverol, Y., *et al.* PRIDE Inspector Toolsuite: Moving Toward a Universal Visualization Tool for Proteomics Data Standard Formats and Quality Assessment of ProteomeXchange Datasets. *Mol Cell Proteomics* 2016;15(1):305-317.
- Pfeuffer, J., *et al.* OpenMS - A platform for reproducible analysis of mass spectrometry data. *Journal of biotechnology* 2017;261:142-148.
- Pluskal, T., *et al.* MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics* 2010;11:395.
- Searle, B.C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* 2010;10(6):1265-1269.
- Slotta, D.J., McFarland, M.A. and Markey, S.P. MassSieve: panning MS/MS peptide data for proteins. *Proteomics* 2010;10(16):3035-3039.
- Sturm, M. and Kohlbacher, O. TOPPView: an open-source viewer for mass spectrometry data. *Journal of proteome research* 2009;8(7):3760-3763.

- Tanaka, S., *et al.* Mass++: A Visualization and Analysis Tool for Mass Spectrometry. *Journal of proteome research* 2014.
- Tsagkrasoulis, D., *et al.* 'Brukin2D': a 2D visualization and comparison tool for LC-MS data. *BMC bioinformatics* 2009;10 Suppl 6:S12.
- Tyanova, S., *et al.* Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* 2015;15(8):1453-1456.
- Winkler, R. MASSyPup--an 'out of the box' solution for the analysis of mass spectrometry data. *Journal of mass spectrometry : JMS* 2014;49(1):37-42.
- Yu, K. and Salomon, A.R. HTAPP: high-throughput autonomous proteomic pipeline. *Proteomics* 2010;10(11):2113-2122.
- Zhang, W., *et al.* LFQuant: a label-free fast quantitative analysis tool for high-resolution LC-MS/MS proteomics data. *Proteomics* 2012;12(23-24):3475-3484.