

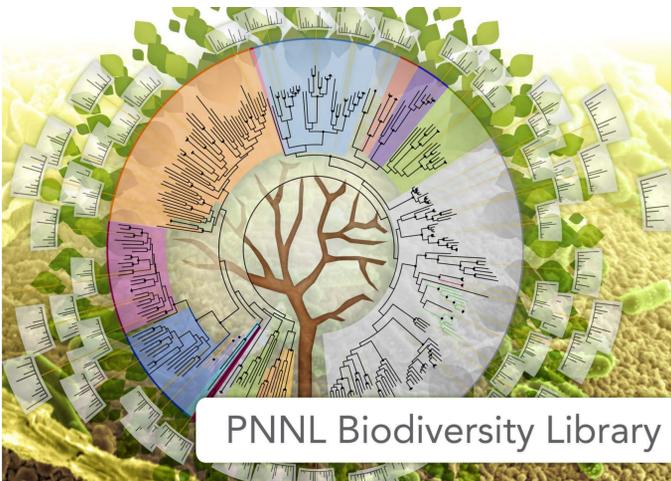
Visual and Intuitive Access to Repository Data

Grant M Fujimoto, Michael Degan, Christopher O Overall, Lillian Ryadinskiy, Christopher Wilkins, Joon-Yong Lee, Richard D Smith, Samuel H Payne



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965



PNNL Biodiversity Library

Skyline Plugin

- PNNL Biodiversity Plugin:
 - C# version 4.5.1, WPF
 - SQLite database backend
- Proteomics data is layered on KEGG pathway images to give users a biologically relevant visual interface to interpret and manipulate data
- Proteins can be selected and exported to Skyline.
- A BiblioSpec spectral library is downloaded and automatically inserted into the active Skyline instance.
- Executes stand-alone or within Skyline

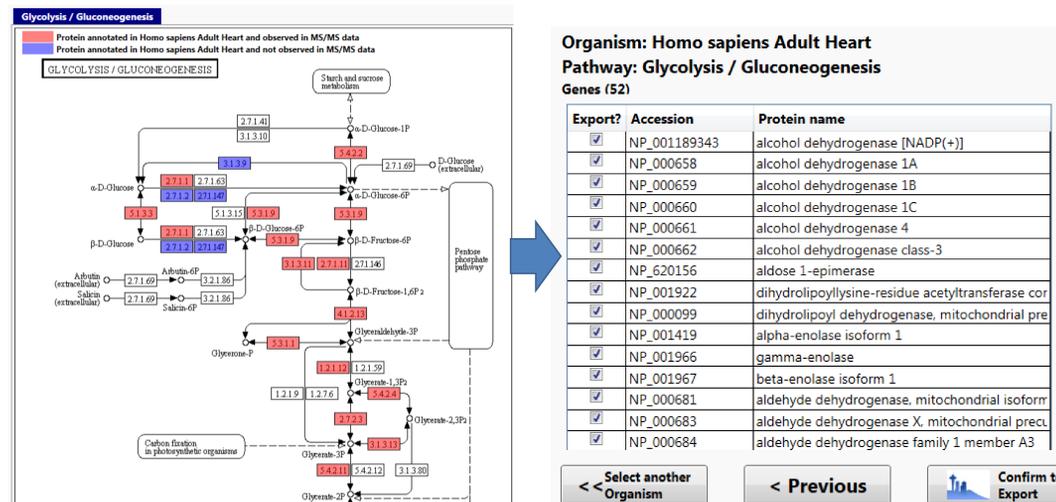


Figure 2 – Plugin Workflow: KEGG ortholog boxes with MS/MS data for are colored red while others are colored blue. All selected orthologs are mapped to proteins to be exported to Skyline.

Overview

- Release of the PNNL Biodiversity Plugin for Skyline which serves as a visual access point to PNNL's mass spec data repository.
- Announce publication and deposition in open repositories of extensive proteomics data for 112 organisms including Human which has 30 separate tissues.
- Proteome coverage per organism averages 80% of functionally annotated genes in KEGG pathways.

Introduction

- A diverse set of organisms has been run by a variety of groups around the world, but they were not analyzed or processed in a standard way
- Publicly available data needs to be presented visually for different uses by various communities
- Few tools exist for viewing data within the familiar context of biological processes and pathways.

Library Data

- Human proteomics data (Kim et al. Nature 2015)
 - 30 adult tissues
- Global proteomics data for 112 bacteria and archaea
 - Raw mass spectra (vendor and mzML formats)
 - Peptide identifications from MS-GF+ (mzid)
 - Condensed spectral library (bibliospec)
 - Proteins mapped to KEGG pathways
 - On average, 2100 proteins per organism
- Data uploaded to ProteomeXchange via MassIVE
- MSV000079053
- PXD001860
- 13 TB of data

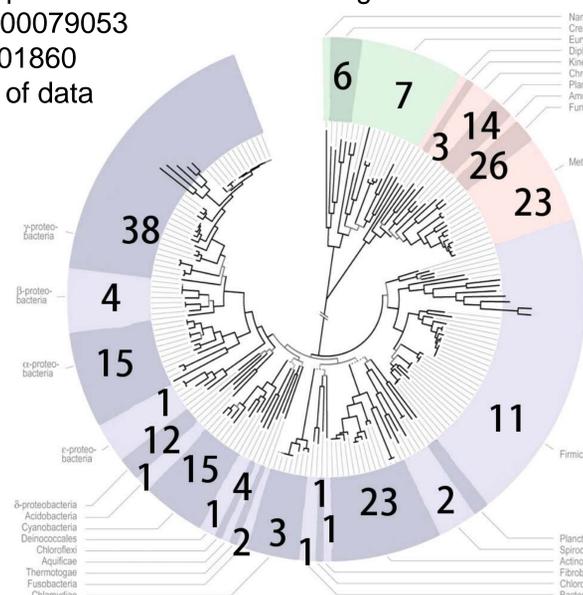


Figure 1 – Diversity. Organisms in the Library are overlaid onto the Tree of Life, showing the breadth of coverage for any particular taxa.

Library Visualization

- Identified peptides are linked to their highest scoring MS/MS scan with relevant peaks annotated.
- Enables higher quality SRM/MRM design.

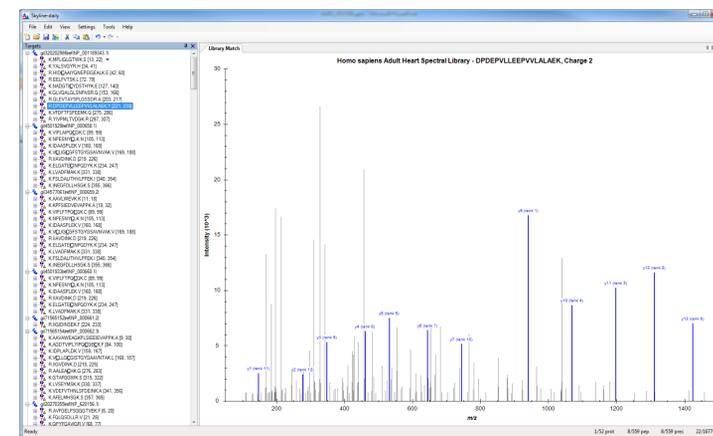


Figure 3 – Skyline All selected proteins are automatically imported to the open Skyline document along with the relevant spectral library.

Conclusion

- The PNNL Biodiversity Plugin is a useful tool for streamlining and simplifying the design of SRM experiments.
- Freely available at omics.pnnl.gov/software or at the Skyline Tool Store.

Acknowledgements

This work was funded by NIH National Institute of General Medical Sciences (GM103493) and an Early Career Award from the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER). Samples were analyzed DOE/BER under the support of NIH National Institute of General Medical Sciences (GM103493), and the Pan-omics program and performed in the Environmental Molecular Sciences Laboratory, a DOE OBER national scientific user facility on the PNNL campus. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under contract DE-AC05-76RL01830.