Scientific Workflows for Proteomics Data Analysis

Yassene Mohammed

University Leiden Medical Center
University of Victoria Genome BC Proteomics Centre
Outline

• PeptidePicker: A Scientific Workflow with a Web Interface for Peptide Selection in Targeted MRM Proteomics

• Other Workflows We Developed

• Taverna Plugin for Proteomics
PeptidePicker

- Designing an MRM assay starts by choosing the most appropriate peptides to represent the target protein

- doi:10.1016/j.jprot.2014.04.018
Determining Most Appropriate Peptides for MRM

• Integrate available information from ExPASy, UniProt, PeptideAtlas, GPMDB, dbSNP, PRIDE and user knowledge to produce weighted list
Web Interface

- Java Servlet
- http://mrmpeptidepicker.proteincentre.com
PeptidePicker Usage (May 1-July 24)
Exploiting PeptidePicker

• Example Chr21:
  – Finished within 30.1h
  – 254 proteins processed/5 failed
  – 500 UniProtKB queries and downloads
  – 6000 ExPASy requests
  – >3500 GPMDB queries
  – >3500 PeptideAtlas queries
  – One copy-paste and one click

• Scalability for 1k and more entries
Other Workflows We Developed

- Cloud Computing
- Spectral Library Generation
- Data Integration and Visualization
Cloud Computing

- Our implementation outperformed MPI and Hadoop implementations
- Cloud is any machine we can access
- Decomposing XML input and recomposing XML output
- Deploy software on the fly

http://pubs.acs.org/doi/abs/10.1021/pr300561q
Spectral Library Generation (in Zebrafish)

• More than 700 datasets, total size of almost 300 GB, few very large files of 3 GB and more

http://pubs.acs.org/doi/abs/10.1021/pr4010585
Taverna Plugin for Proteomics
Taverna Plugin for Proteomics
Acknowledgments

• Leiden University Medical Center
  Andre Deelder
  Magnus Palmblad

• University of Victoria
  Dominik Domanski
  Angela Jackson
  Derek Smith
  Christoph Borchers

• Hogeschool Leiden
  Sidney Niccolson
  Matthijs Kuik
Many Thanks!