Proteome Informatics Research Group

**Proteome Informatics**

**Research Group**

**Abstract**

Accurate and concise reporting of protein identification data that result from mass spectrometry-based proteomic workflows is a key bioinformatic challenge. One particular complication, referred to as the protein inference problem, results from the loss of peptide to protein mapping caused by the enzymatic digestion of protein mixtures. Indeed, the Paris Guidelines, which represent the proteomics community’s efforts to devise a standard methodology for reporting protein identification data, puts particular emphasis on reporting a minimal set of proteins that account for all of the identified peptides. In the present work, the newly formalized Proteome Informatics Research Group (iPRG) presents the results of a study to assess the quality and consistency of protein inference analysis across the ABRF community.

**Introduction**

Primary goal: What is the current state of protein reporting?

- Does the proteomics community still have a problem with reporting excessively long protein lists due to poor protein inference?
- Is ambiguity among multiple non-differentiable accessions being properly reported in proteo groups?
- Assess the consistency of protein identification analysis across users starting from the same mass spectra.

Secondary goal: Establish a benchmark reference.

- Enable users of software to test their lab’s methods.
- Provide a frame of reference for future software development.

**Methods**

- A mouse liver differential expression experiment with trypsin digestion, MMTS alkylation, and iTRAQ® reagent labeling was separated into 13 cation exchange mixtures. The iPRG thanks Renee Robinson (Harvard University) for serving as our anonymizer.
- The results were submitted to a 2008 iPRG survey producing 29 files and 432 spectra.
- The data were provided in raw, SWATH, TMT, and mzML, and mzXML format. Use of the provided database (MD, Dec 1, 2007 version, 53,826 proteins + 74 contaminants) or a derivative of it was required.
- Users starting from the same mass spectra. The Paris Guidelines represent the proteomics community’s efforts to devise a standard methodology for reporting protein identification data. Indeed, the Paris Guidelines, which represent the proteomics community’s efforts to devise a standard methodology for reporting protein identification data, puts particular emphasis on reporting a minimal set of proteins that account for all of the identified peptides. In the present work, the newly formalized Proteome Informatics Research Group (iPRG) presents the results of a study to assess the quality and consistency of protein inference analysis across the ABRF community.

**Figure 1: Results Overview**

**Figure 2: Number of Protein Groups by Cluster**

**Figure 3: Cluster 101 (17 groups, 14 expected)**

**Figure 4: Reporting Protein Ambiguity**

**Conclusions**

- Virtually all respondents used some kind of protein inference tool.
- No gross inflation in number of reported proteins was seen in any submission, although in some cases mistakes were made.
- Some submissions did not properly report accession ambiguity (3/18).
- Not enough submissions to draw conclusions across the proteomics community.
- More submissions are needed! Please email iPRG2008@gmail.com if interested.

**Acknowledgments**

The iPRG thanks Renee Robinson (Harvard University) for serving as our anonymizer.