SILAC Quantitation to a Depth of 3000 Proteins from a Double Knockout GSK-3 Line of Mouse Embryonic Stem Cells

Chris Hughes[1], Brad Doble[2], Lei Xin[3], Clark Chen[3], Baozhen Shan[3], Bin Ma[3], Gilles Lajoie[1]

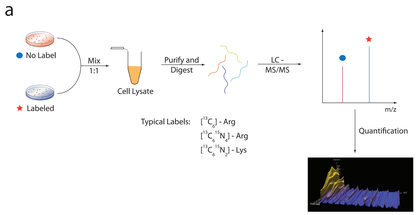
1 - University of Western Ontario, London, Canada 2 - McMaster University, Hamilton, Canada 3 - Bioinformatic Solutions Inc., Waterloo, Canada

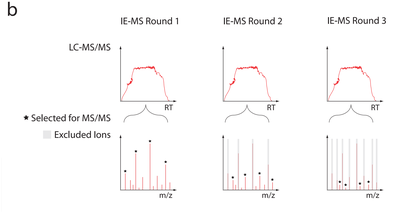
Introduction

The use of stable isotope labeling of amino acids in culture (SILAC)[1] in quantitative proteomics research is becoming increasingly common. SILAC provides a powerful means towards extracting more meaningful information from a proteomics dataset. In experiments involving the knock-down or knock-out of gene expression, measuring the large-scale effects on protein expression levels remains a challenge. In order to completely understand the resulting changes, quantitative information on a large number of proteins needs to be obtained. Metabolically labeling proteins with SILAC enables relative quantitation for a large number of candidates.

In small scale experiments dealing with the analysis of relatively few proteins, quantification can be performed manually. However, the analysis of experiments which result in the identification of thousands of proteins continues to be difficult, due in part to complications resulting from large datasets reaching several hundred gigabytes size. Up until now, the Trans-Proteomic Pipeline[2] was the only platform for large-scale identification and quantitation of proteomics data that is independent of file format. Here we will demonstrate quantification of a mouse embryonic stem cell dataset containing ~450 gigabytes of raw data done effectively with PEAKS 5 software[3].

Methods





Mouse embryonic stem cells were grown on gelatin coated plates in DMEM based media supplemented with KOSR and LIF. Cell lysates were mixed 1:1 based on protein concentration and separated by 1D SDS-Gel followed by in-gel digestion with trypsin. Peptides were analyzed on a Q-ToF Ultima from Waters using data-dependent scans with iterative exclusion[4]. Separation was performed on a nano-Acquity LC equipped with a 15cm x 75um C18 reverse phase column using a 150 minute solvent gradient.

Results

The complete dataset consists of 185 raw datafiles combining to a size of ~450 gigabytes. There are total of 512,455 MS/MS spectra within these raw files. Each raw file was searched using PEAKS 5.0 and X!Tandem. The dataset was also searched using SpectrumMill with peak list files as input. Valid hits were those with at least 2 unique peptide identifications and scores above set thresholds for each specific engine. Full quantitation for SILAC labeled peptides was carried out using PEAKS 5 and the Trans-Proteomic Pipeline (TPP) with XPRESS function.

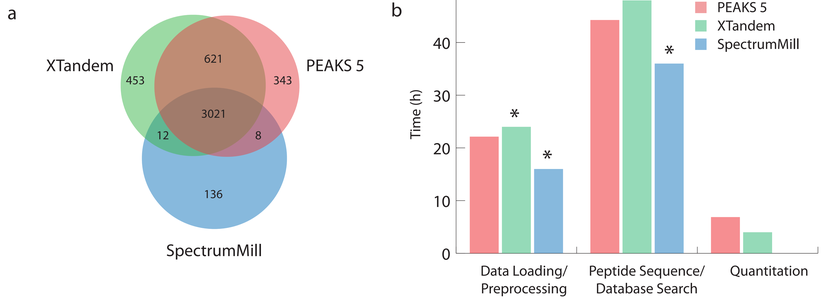


Figure 1 -Search engine results obtained from this study. (a) Depicts the protein identification overlap between the three database search engines employed. (b) Processing times for different stages of analysis based upon analysis of the entire dataset. * denotes processes where user intervention is required. Dataset comparison was performed using LabKey Software Version 9.1.

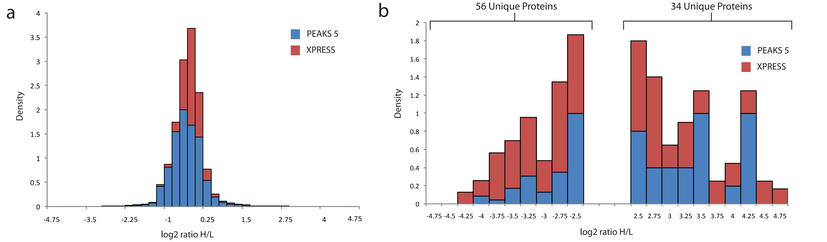


Figure 2 - Quantitation results obtained using PEAKS 5 and XPRESS algorithms. Log ratio values are binned in groups of 0.25 difference in size. (a) Depicts the distribution of protein ratios obtained from both quantitation softwares for the whole dataset. (b) Represents the number distribution of proteins which were up or down-regulated greater that 2.5 fold.

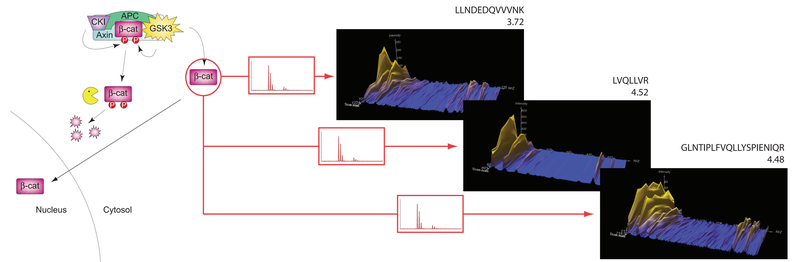


Figure 3 - Effects of GSK-3 knockdown on peptides belonging to beta-catenin. All three spectra depict extracted ion chromatograms representative of the peptide ratios quantitfied with PEAKS 5.0. Heavy peptides are derived from the wild-type cells, whereas light peptides are from the double-knockout cell line. GSK-3 pathway adapted from Jope, R.S. et al. (2004) *Trends in Biochem. Sci.*

Effective knockout of GSK-3 expression is validated by the disappearance of any GSK-3 peptides in the mutant cells as depicted by the SILAC ratios. Strong up-regulation of the beta-catenin protein shows a decrease in degradation signaled by GSK-3. The majority of the ratios obtained were centered around a H:L ratio of -0.50, indicating that most identified proteins did not significantly alter in abundance as a result of the knockout. However, there were 56 proteins which were up-regulated in the knockout cells which may be a direct or indirect result of a reduction in GSK-3 levels.

Conclusion

The use of SILAC permits the monitoring of pathways on a global level. However, without the ability to quantitate on large sample sets, there is a limitation on how much information can be extracted. As is shown here, PEAKS permits large scale analysis in a streamlined fashion, comparable to that of the TPP. The dataset has recently been expanded to >800000 MS/MS spectra from ~810 gigabytes of data. Further analysis on the global phosphorylation states of detected proteins using PEAKS is ongoing in an attempt to map modification changes as a result of the GSK-3 knockout.

References & Acknowledgements

1 - Ong, S.E., Kratchmarova, I., and Mann, M. (2003) *J Proteome Res* 2, 173 - 181
2 - Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002) *Anal Chem* 74(20), 5383 - 5392
3 - Ma, B. *et al.* (2003) *Rapid Comm Mass Spec* 17(20), 2337 - 2342
4 - Bendall, S.C. *et al.* (2009) *Mol. Cell. Proteomics* 8(3), 421 - 432

The authors wish to thank Agilent Technologies Inc. for granting an award to C.H. for travel to ASMS.