New Cross Experiment Quantification Tool for SILAC/TMT: PEAKS Q

Lei Xin, Zia Rahman, Hao Lin, Baozhen Shan
Bioinformatics Solutions Inc., Waterloo, ON, Canada

PEAKS®
Complete Solution for Proteomics

Purpose

1. To develop a software tool for labelling quantification using the MultiNotch MS3 approach
2. To support increasing multiplexing by combining experiments for large-scale protein quantification design

Introduction

Protein quantification has become the main research interest in proteomic studies. Among the quantitative methods, precursor ion labelled quantification method such as SILAC is most often used due to its accuracy. Recently, reporter ion labelled quantification methods such as TMT/iTRAQ have gained a lot of interest due to increased accuracy using the MutliNotch approach [1]. However cross experiment data analysis always pose a challenge to the researchers. In this study, we evaluated the performance of PEAKS Q software for cross experiment MS data analysis. This new quantification module of PEAKS can handle labelled quantification at both the precursor level and reporter ion level.

Method

One set of data is used for each type labeling method. For SILAC study, we used a publicly available dataset (PXD000151) containing 3 replicates [2]. Heavy isotopically labelled forms of Arginine and Lysine (Arg10-13C615N4, Lys8-13C615N2) were used for SILAC labelling. Equal amounts of unlabelled and labelled human Embryonic Stem Cell (hESC) proteins under different conditions were mixed together followed by LC-MS/MS run on a Thermo Q Exactive. The raw data analysis was carried out using PEAKS DB for identification followed by quantitative analysis using PEAKS Q.

For the TMT study, we analyzed a dataset containing 3 replicate runs that was generated using the MultiNotch MS3 method with TMT8-plex isobaric tags, acquired on Orbitrap Fusion. In the first four channels 126, 127N, 127C and 128, digested E. Coli proteins were mixed with ratio 10:5:2:1. In the remaining four channels 129N, 129C, 130 and 131, digested E. Coli proteins were mixed with ratio 1:2:5:10 and spiked in the Hela digest. The raw data was analyzed by PEAKS DB for identification followed by quantitative analysis using PEAKS Q.

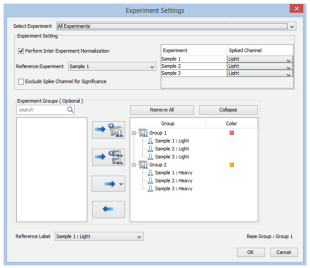

Figure 1. Cross experiment analysis for SILAC data

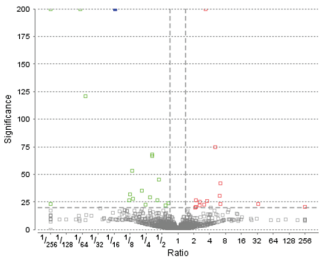

Figure 2. Protein volcano plot for SILAC analysis

SILAC Analysis

The protein identification is performed using PEAKS DB against Uniprot human database using decoy database searching for False Discovery analysis. The search yields 2703 distinct hESC proteins groups at 1% FDR. Total 2406 distinct proteins groups (>89%) were quantified by PEAKS Q module that uses MS1 feature based quantification for SILAC. The protein ratios exhibit a clustering around 1:1 as expected. The significances of differential expressions were analysed using PEAKS Q method which is an extension of Significance B method introduced in paper [3]. We have found that only 33 protein groups exhibit significantly differential expression (>20 PEAKS Q significance score).

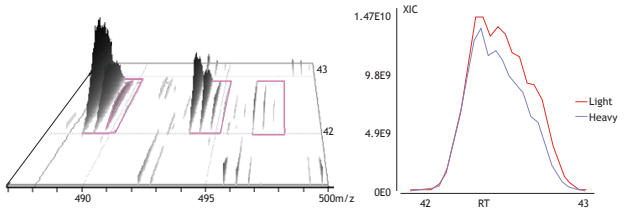

Figure 3. Auto-correction for R-to-P conversion

Although extensive measurements were taken during sample preparation to avoid Arginine-to-Proline (R-to-P) conversion by reduction of the labelled Arginine, PEAKS Q was still able to detect the features related to the R-to-P conversion and intensity were automatically corrected for accurate quantitative analysis.

TMT Analysis

The protein identification is performed using PEAKS DB against Uniprot database with E. Coli and human proteins using decoy database searching for FDR analysis. The identification search results in 1770 distinct protein groups with at least 1 unique peptide at 1% FDR. Total 1763 protein groups out of these 1770 protein groups (>99.5%) were quantified by PEAKS Q. The validation tools (protein profile heatmap, intensity boxplot, etc.) included with PEAKS Q allowed us to confirm that the calculated protein ratios show a similar trend as the expected ratios. By adjusting quantification p-value to FDR using Benjamin-Hochberg method, with 1% FDR 69 proteins show significant changes. Among them 67 are E.Coli proteins.

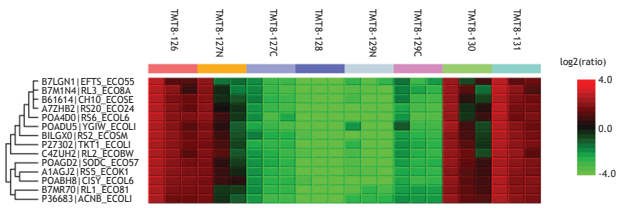

Figure 4. Protein profile heat map for TMT analysis

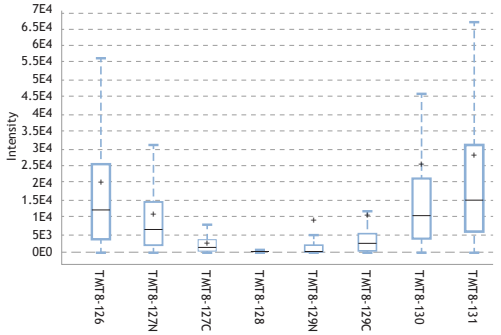

Figure 5. Intensity boxplot for TMT channels

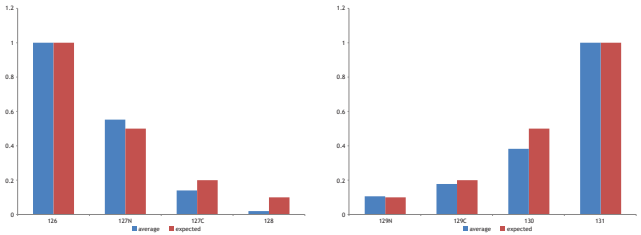

Figure 6. Average ratio of E. Coli proteins compared to the expected ratio

Conclusion

PEAKS Q as a cross experiment quantification tool makes it easy to handle MS data from multi-run TMT/SILAC experiments. Taking advantage of the improved feature detection algorithm and significance evaluation algorithm, the software can detect protein expression change with high sensitivity. All the statistical charts and data link functions make it very convenient to validate quantification results with raw data.

Reference

[1] G.C. McAlister, D.P. Nusinow et al. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. Anal. Chem., 86(14):7150-7158, 2014.

[2] A.R. Liberski, M.N. Al-Noubi et al. Adaptation of a Commonly Used, Chemically Defined Medium for Human Embryonic Stem Cells to Stable Isotope Labeling with Amino Acids in Cell Culture. J. Proteome Res., 12(7):3233-3245, 2013.

[3] J. Cox and M. Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology, 26(12):1367-1372, 2008.