Systematic Assessment of the **Reproducibility** of Relative Quantification Based on LC-MS with Replicates

Baozhen Shan[1], Weiwu Chen[1], Bin Ma[2]
[1] Bioinformatics Solutions Inc, Waterloo, ON
[2] University of Waterloo, Waterloon, ON

Bioinformatics Solutions Inc.

Overview

Purpose: To develop a method to assess and display the reproducibility of protein quantification based on LC-MS with replicate analysis.

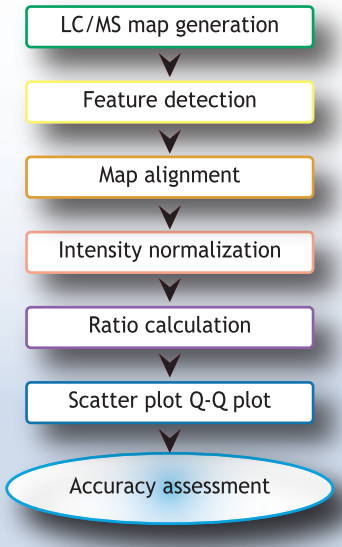
Methods: Technical replicates are used to assess the reproducibility of LC-MS for differential analysis. Statistical analysis. Results are visualized by biplots.

Results: Reproducibility assessment helps to evaluate both the LC-MS experiment and the quantification algorithm.

Introduction

Since the quantification of a protein's expression level relies exclusively on the mass spectrometric intensity of each peptide ion, the reproducibility of the intensity measurement remains the essential factor for reliable LC-MS based proteomics. Knowing the reproducibility of the system provides more confident expression ratios, which are used as criteria for differential analysis (Figure 1). An understanding of the system of reproducibility can be achieved through the application of quality control processes in conjunction with a reproducibility assessment, which will minimize the number of false positives and negatives in differential analyses. One approach that ensures greater experimental success is the incorporation of replicates.

Figure 1. Flow Chart process for reproducibility assessment



Methods

Technical replicates were used to assess the reproducibility of relative quantification uniquely at three levels:

Data level - Peptide features from each of the replicates were detected and aligned between the replicates. The data reproducibility was analyzed by paired T-test and correlation analysis on the intensities of paired features. The boxplot diagram was used to represent the distribution of the logarithmic ratios of feature pairs. Venn diagrams of peptide features between replicates were also used to display the reproducibility.

Identification level – Protein identification was conducted on each replicate. The reproducibility of the identified peptides and proteins between replicates were assessed by paired T-test and correlation analysis. The boxplot diagram was used to represent the distribution of the logarithmic ratios of identified feature pairs.

Quantification level - The relative peptide/protein abundance was computed based on the intensity of each peptide ion. The reproducibility of peptide and protein quantification was assessed by paired T-test and/or ANOVA.

Figure 2. Data organization of replicates in PEAKS®

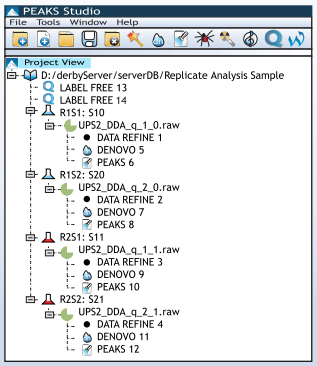
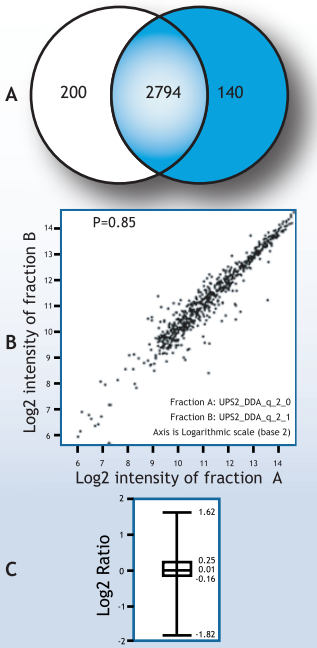


Figure 3. Scatter plot of peptide feature pairs

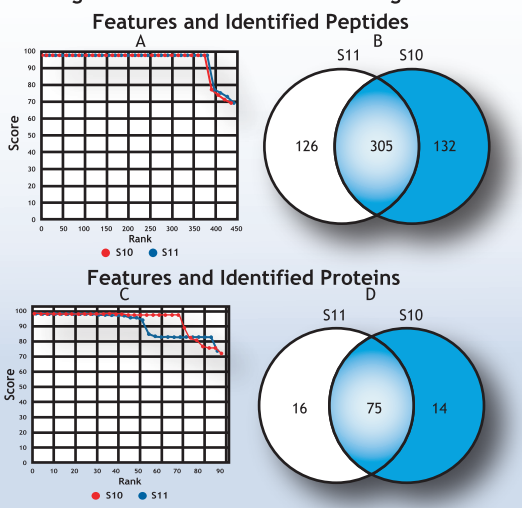


Results

This method was implemented into the PEAKS® Studio software and tested by a standard protein mixture. The standard mixture was reduced and alkylated by iodoacetamide, then digested by trypsin overnight. MS data was obtained from a Waters Q-TOF Premier. Two replicate experiments were performed. MS raw data was analyzed with the PEAKS® de novo sequencing, database search and label-free quantification algorithms (Figure 2). S10 and S11 are two replicates of one sample and S20 and S21 are replicates of another sample. The results of reproducibility assessment at data level are shown in Figure 3. The number of peptide features in sample S10 and Sample 11 are 2994 and 2934, and the number of common features is 2794 (Figure 3A). The scatterplot (Figure 3B) and boxplot of logarithm of the ratio's of intensities (Figure 3C) of common peptide features showed a good data reproducibility. The Pearson correlation value is 85% for common features.

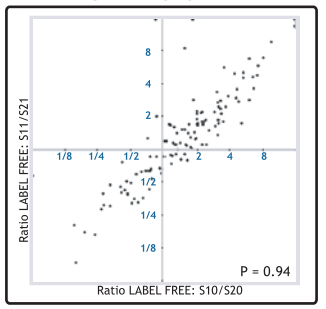
The results of reproducibility assessment at the identification level are shown in Figure 4. The total number of identified feature pairs (at least one feature in the feature pair is identified) in S10 and S11 is 563. The corresponding Pearson correlation value is 92%. Figures 4A and 4B are the score chart and Venn diagram for identified peptides and 4C, and 4D for identified proteins S10 and S11.

Figure 4. Score chart and Venn diagrams



Unique peptides are used for quantification in PEAKS®. 84 of 563 feature pairs were unique and used for protein quantification. The scatter plot of peptide relative abundance in two replicates is shown in Figure 5. The corresponding Pearson correlation value is 91%. Figure 5 shows a good reproducibility of label-free quantification.

Figure 5. Scatter plot of peptide ratios



Conclusions

This method proves to efficiently assess and display the reproducibility at three levels, which helps to evaluate both the LC-MS experiment and the quantification algorithm.

References

Kim Y.J. et al. Reproducibility Assessment of Relative Quantification Strategies for LC–MS Based Proteomics, Anal. Chem., 79. 5651 (2007).

Acknowledgement

We thank Dr. Gilles A. Lajoie for the mass spectrometry experiment.