Blocking Design

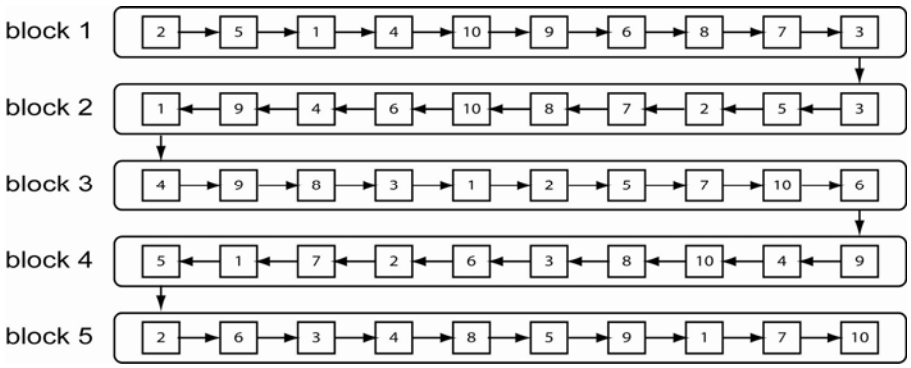

Figure 1. Blocking design. Each small box refers to an LC-MS dataset with the number inside it indicating the time point. Each of the larger boxes refers to a block that contains a complete time series of LC-MS datasets. Within each block, the order of the LC-MS analyses is random. The arrows indicate the time order that the LC-MS analyses were performed.

Abundance Normalization

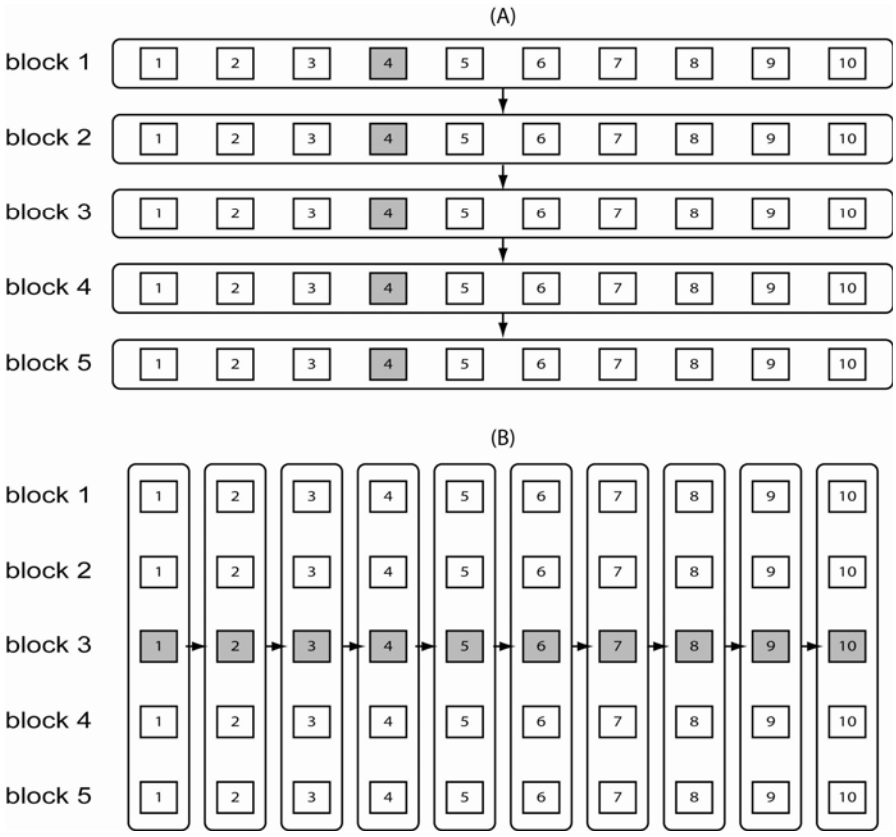

Figure 2. (A) Selection of reference LC-MS analyses (shaded) for abundance normalization of the time series data. Normalization was performed block by block. The reference datasets across all of the blocks correspond to the same time point. (B) Selection of reference LC-MS analyses (shaded) for abundance normalization of the technical replicate data. Normalization was performed time point by time point. The reference LC-MS datasets corresponding to the different time-points are all within the same LC-MS block.

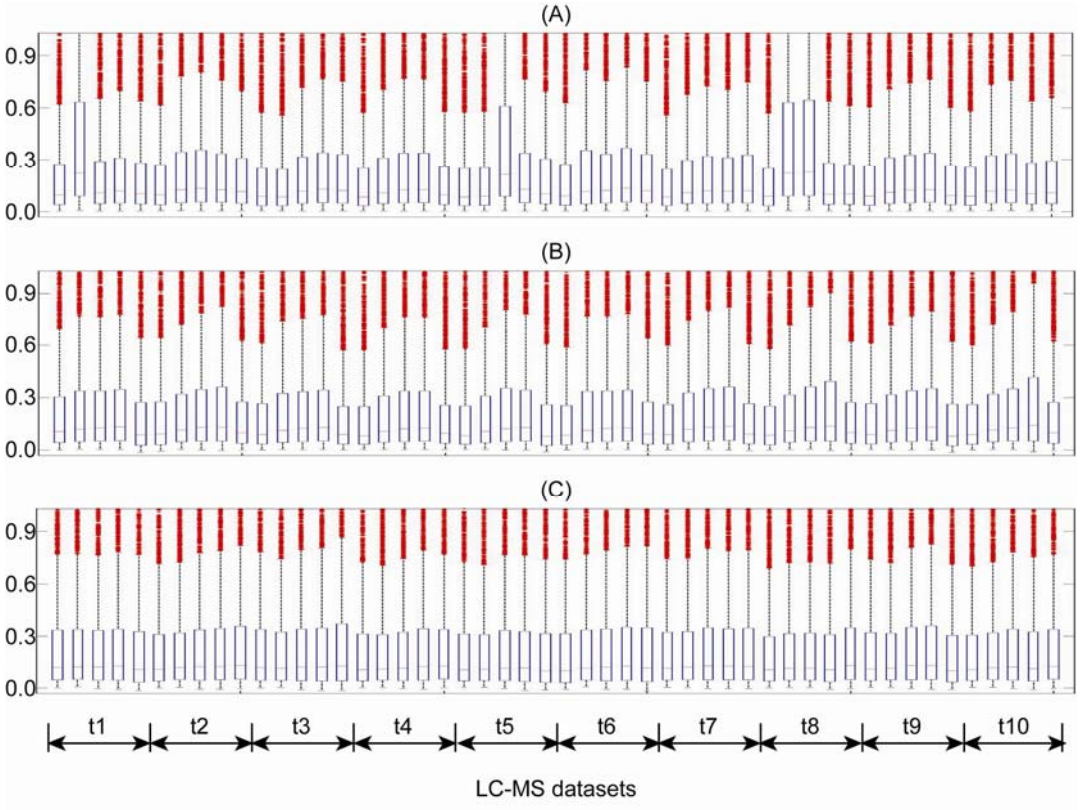

Figure 3. Boxplots of all of the LC-MS datasets before and after abundance normalization. All the datasets are grouped based on time and the 5 technical replicate datasets corresponding to the same time point are placed together. t1-t10 denotes time points. The 1^{th} and 99^{th} percentile of the abundance values were chosen as the lower and the upper abundance limits for calculating the regression lines for the time series and technical replicate data, respectively. (A) Boxplots before abundance normalization. (B) Boxplots after normalization of the time series data during the first iteration. While normalization clearly brought the centers of the abundance distribution substantially closer to each other for all LC-MS datasets within each block, the median lines of all technical replicate datasets that correspond to the same time point still deviate from each other. As a result, the technical replicate datasets also needed to be normalized. (C) Boxplots after normalization of the technical replicates data during the first iteration. The systematic shift that existed between pairs of technical replicate datasets corresponding to the same time point has been reduced.

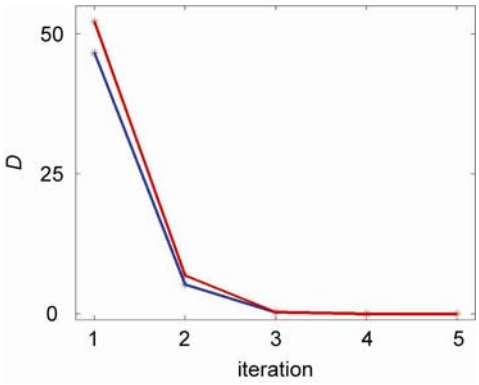

Figure 4. Convergence of the abundance normalization. The normalization process converged as the total change D in the regression lines between adjacent iterations approached zero. The convergence of D that corresponds to the normalization of the time series data and the convergence of D that corresponds to the normalization of the technical replicate data are shown in red and blue, respectively.

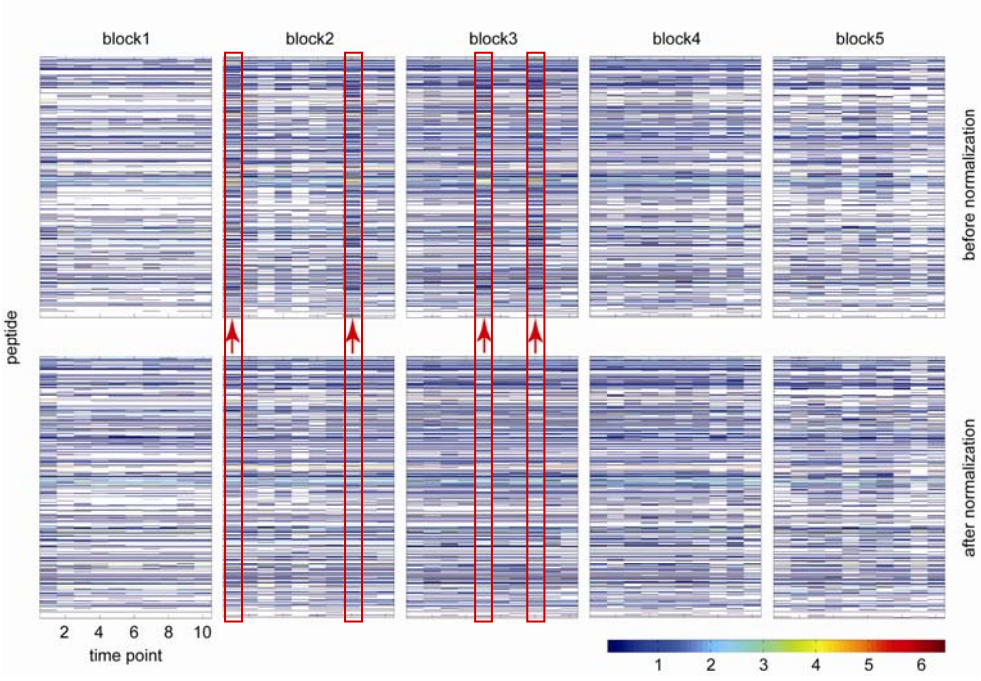

Figure 5. Heat-map of abundance values in the filtered peptide list before and after abundance normalization. Each sub-figure corresponds to one block. Datasets (indicated by red rectangles) in which peptide abundance values were prominently higher than the peptide abundance values in the rest of the datasets appear in blocks 2 and 3. (Top row) Heatmap before abundance normalization. (Bottom row) Heatmap after abundance

normalization. The prominent differences in abundance values among datasets have been reduced, which corroborates the boxplot results shown in Figure 3C.

Identification and quantitation of peptides using the AMT [1] tag approach

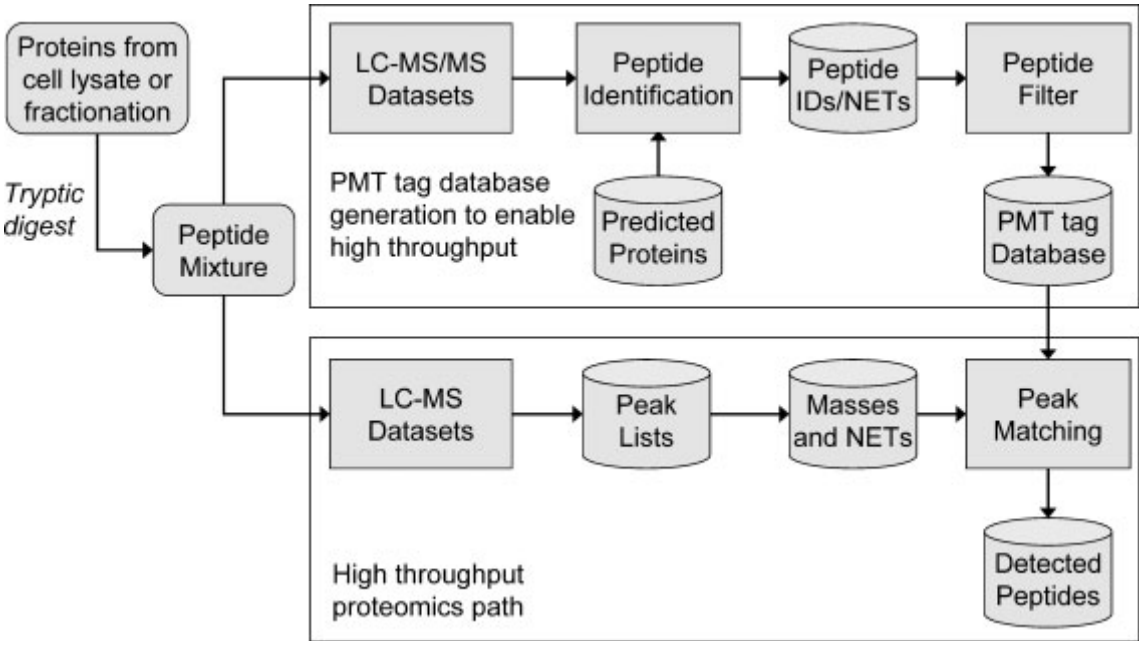

Figure 6. AMT tag process.

The AMT tag process (Figure 6) can be generally divided into two steps: (1) Generation of the putative Mass and Elution Time (PMT) tag database. Proteins from cell lysates are digested and analyzed by LC-MS/MS. Peptides are then identified, and their elution time is predicted and normalized. These peptides are finally deposited into the PMT tag database, with each peptide characterized by its sequence, theoretical mass, and normalized elution time (NET). (2) Quantitation of peptides using LC-MS analysis. An eluted peptide is generally scanned consecutively multiple times in the MS analysis and the sum of the peptide intensity in all the scans are calculated as the total abundance of this peptide (Figure 7). (3). Identification of peptides. The list of peptides from the LC-MS analysis are featured as a (mass, elution time, and abundance) triple. The mass and elution time are used to identify the peptides by looking for peptides that have similar mass and elution time in the PMT tag database. Those peptides that have close matches are identified, and those that do not are unidentified and will be missing in the final list of peptides that are identified and quantified. For a more comprehensive description of the peptide identification and quantitation using the AMT tag approach, please refer to reference 1.

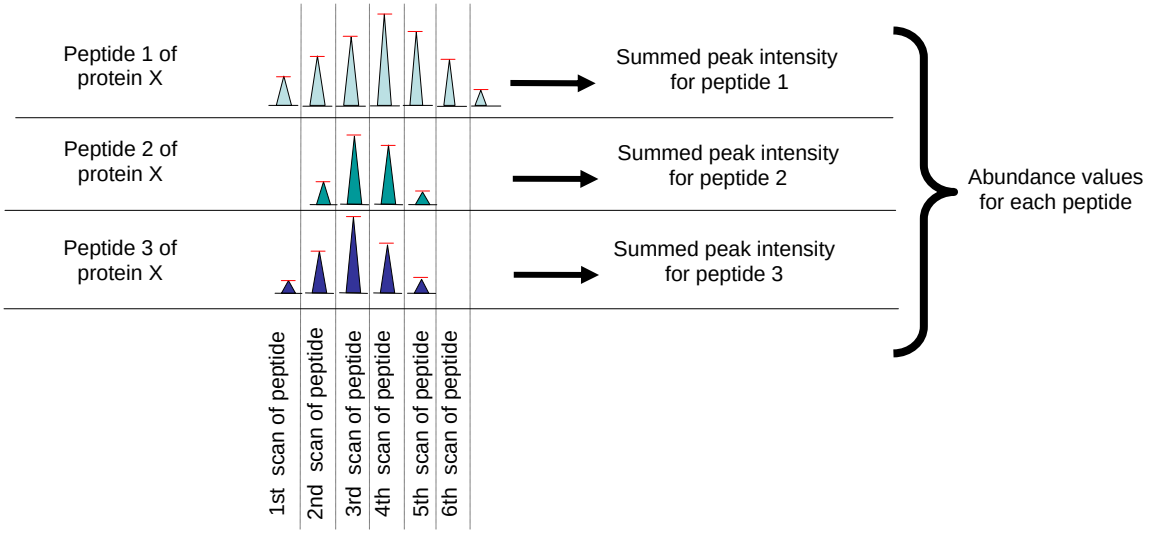

Figure 7. Quantitation of peptides using an LC-MS analysis

References

1. Zimmer, J. S.; Monroe, M. E.; Qian, W. J.; Smith, R. D., Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev* **2006,** 25, (3), 450-82.