



## M006.1P – Proteomic Data Statistical Processing

The peak intensity values (i.e., abundances) for the final peptide identifications are processed in a series of steps using MatLab® R2010b that included quality control, normalization, protein quantification, and comparative statistical analyses. Peptide abundances are transformed to the log<sub>10</sub> scale. Missing data values are not imputed. Quality control processing is performed to identify and remove peptides with an insufficient amount of data across the set of samples (1), and identify and remove LC-MS runs that showed significant deviation from the standard behavior of all LC-MS analyses (2). LC-MS runs are identified as an outlier at a significance level of 0.0001. Peptides are normalized using a statistical procedure for the analysis of proteomic normalization strategies ([SPANS](#)) that identifies the peptide selection method and data scaling factor which introduces the least amount of bias into the dataset (3). The peptide abundance values are normalized across the technical replicates using a rank invariant subset of peptides (p-value threshold of 0.1) followed by median absolute deviation centering of the data. Normalized log<sub>10</sub> abundance values are averaged across the technical replicates within each biological sample. Proteins are quantified using a standard R-Rollup method (4) using the most abundant reference peptide, after filtering the peptides that were redundant, had low data content, or were outside the dominant significance pattern. Comparative statistical analyses of time-matched mock with infected samples are done using a Dunnett adjusted t-test to assess differences in protein average abundance, and a G-test to assess associations among factors due to the presence/absence of response.

### References:

1. Webb-Robertson, B.J., et al., Combined statistical analyses of peptide intensities and peptide occurrences improves identification of significant peptides

Contact us at <https://www.systemsvirology.org/project/home/begin.view>



- from MS-based proteomics data. *J Proteome Res*, 2010. 9(11): p. 5748-56.
2. Matzke, M.M., et al., Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics*, 2011. 27(20): p. 2866-72.
  3. Webb-Robertson, B.J., et al., A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. *Proteomics*, 2011. 11(24): p. 4736-41.
  4. Polpitiya, A. D., W. J. Qian, N. Jaitly, V. A. Petyuk, J. N. Adkins, D. G. Camp, G. A. Anderson, and R. D. Smith. 2008. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24:1556-8.

