An Integration Methodology for Combining Results from Multiple Search Algorithms

Daehee Hwnag

Institute for Systems Biology. WA, U.S.A.

Database-searching algorithms generally identify only a fraction of the spectra acquired from LC-MS/MS analysis, due to several issues. These issues include noisy spectra hindering correct identifications and also systematic biases in searching algorithms (favoring identification of particular types of peptides). As a result, there often exist small overlaps between search results when multiple search algorithms were applied to the same dataset.

To resolve this problem, here we present an integration methodology that can combine effectively results from multiple search algorithms using statistical meta-analysis. Our methodology first weights individual search methods based on their contributions to maximizing the number of potential correct identifications (i.e., maximizing statistical power). Then, statistical significances (p-values) of search results from search methods are combined using such weights based on several methods (Fisher's method or Stouffer's Z) in meta-analysis. To demonstrate the efficacy of our methodology, we generated a dataset for 7 protein mixtures using LC-MS/MS analysis, and then applied our methodology to combine search results from the following search algorithms: SEQUEST, ProbID, and Comet. The results show that different types of systematic biases from these search algorithms were efficiently handled by the weights. Also, combined p-values from our methodology provided more correct identifications than individual algorithms, when the same significance level cutoff value (e.g., $p = 0.05$) was used.

We are currently extending this methodology to include more search algorithms such as Mascot and X!tendam. Also, the validity of this methodology is being further evaluated by applying it to various datasets from complex samples and several mass spectrometry platforms.