Computational Challenges in Identification and Quantification

ASMS Bioinformatics Interest Group (BIG) Meeting
Tuesday, May 22, 2012, Vancouver, Canada

Brian C. Searle & Marc Kirchner, presiding

Summary

The 2012 BIG meeting focused on recent technological developments and how they affect computational approaches to solving proteomics problems. The goal was to spark discussions on the following two questions:

- Does having highly mass accurate fragment ions substantively change how bioinformaticians should approach peptide identification algorithms?
- Do the quantitative benefits of acquiring fragment ions for all peptides simultaneously and thereby spending more time acquiring quantitatively accurate parent ion intensity data outweigh the benefits of isolating fragment ions for each peptide individually?

While certain instruments have been able to acquire high-resolution fragment ions in proteomics experiments for at least a decade and algorithmic strategies for deconvoluting data-independent acquired fragment ions from multiple peptides have existed for many years, recent advances in instrumentation have started to popularize these methods. We invited two speakers, John Cottrell from Matrix Science Ltd. and Nuno Bandeira from University of California, San Diego, to briefly reflect on each question before opening up the discussion to the audience. The speakers were asked to be as provocative as possible and present opposing views to spark discussion. The intention of the meeting was to introduce these technological advances to bioinformaticians in the audience and inspire them to take different approaches to protein identification and quantification.

Approximately 150 people attended the “standing room only” interest group meeting and we enjoyed a lively discussion spending approximately thirty minutes on each topic. The discussion on high-resolution fragment ions revolved around the roles database searching and de novo sequencing play in protein identification. While high-resolution fragment ions improve de novo sequencing accuracy, it is clear that there are still questions about how to assess statistical significance of the sequences. Database searching can also usually provide more useful information from marginal data. At the same time, de novo sequencing doesn’t require a sequenced genome and can usually interpret more diverse post-translational modifications. While there will likely still be a place for traditional database searching in proteomics workflows, de novo sequencing may play a larger role than it has in the past.

The discussion on data-independent acquisition strategies (MS^E, SWATH, CDA) focused on whether deconvolution could produce comparable quality peptide identification results to the industry standard top-N acquisition strategies that rely on exclusion windows to isolate peptide fragment ions. It was clear from the discussion that prior knowledge of relative retention times and fragment intensities from spectral libraries will play a critical role in the accurate interpretation of this type of data.