

The Informatical Difference between Targeted and Discovery-based Proteomics

Monday, June 10, 2013

Brian C. Searle and Lukas Käll, presiding.

The 2013 Bioinformatics for MS Interest Group Workshop focused on the implications of the renewed interest in alternative analyte fragmentation sampling strategies that are grouped under the term “data independent acquisition” (DIA). Such methods have some nice properties facilitating the downstream informatical analysis, such as increased reproducibility and more precise quantification. However, fragmentation spectra stemming from DIA are often complex in the sense that they may contain fragments from multiple peptides, which challenges the current strategies for peptide identification. Two conflicting ideas of how to process data have been proposed: spectrum-centric queries, where one in resemblance with discovery based searches find the best matching peptides for each spectrum, or peptide-centric queries, where one like in targeted proteomics find the best scoring spectral evidence for each hypothesised peptide.

We invited two speakers, Michael MacCoss from University of Washington and Scott Geromanos from Waters Inc., to briefly reflect on the 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. As this is an area of rapid development, the intention of the meeting was to introduce the audience to the choices that the developers of processing tools will have to make, and what consequence such choices will have. Approximately 200 people attended the interest group meeting and we enjoyed a discussion full of insights.

While the peptide-centric searches appears as a straightforward way to deal with the chimeric nature of DIA spectra, as the identification of a peptide from a spectrum do not affect additional identifications from the same spectrum. The approach also favours reproducible analysis as it guarantees a score for each hypothesised peptide. The spectrum-centric approach on the other hand have other nice properties, such as facilitating the interference-free fragment-ions which facilitates quantification, and that it can profit from existing database search technology.

For some implementations of the spectrum-centric approach it has been noted that an increased number of collected spectra leads to a deteriorated identification rate, due to the increased effect of multiple hypothesis corrections. However, it was argued that sooner or later the community would like to expand the amount of searched peptides allowing for multiple-PTM or *de Novo* searches based on DIA data, and under such conditions the multiple hypothesis corrections will severely affect the sensitivity of peptide-centric searches as well.

It was clear from the discussion that increased understanding of the fragmentation patterns of the particular peptides we aim at identifying can greatly increase our chances of successful interpretation. Also, as we extend our ability to detect lower abundant fragments with DIA, we need to decrease the complexity of our samples by efficient prefractionation, to facilitate identification.