Analysis of mass spectrometry data using sub-spectra and wavelets

Spectra resulting from Surface-Enhanced Laser Desorption/Ionisation (SELDI) mass spectrometry measurements are constructed by combining sub-spectra, each of which are the result of a single firing of the laser responsible for the process of desorption/ionisation. These firings are performed, with several repeats, at different locations of the spot on which the sample is analysed. The final spectrum is then constructed by taking the mean over all these sub-spectra. We find this is sub-optimal in that it is likely to average out peaks from peptides that are present in low abundance or are unevenly distributed over the spot. Furthermore, we find that characteristics such as the amount of noise varies considerably between sub-spectra, indicating experimental factors such as laser inefficiency, but also spatial differences in total protein and matrix material content. This argues for analysing sub-spectra separately and combining results afterwards.

Here, we analyse sub-spectra using methods based on wavelets ([1, 2]), which account for the presence of noise implicitly by assuming its frequency and shape is fundamentally different from that of the signal. In doing so, we gain more control over the process of distinguishing signal from noise. This allows us to detect additional peaks representing low(er) abundance proteins, thereby increasing the sensitivity of the SELDI technique. By analysing results over multiple sub-spectra, we reduce the number of false positives dramatically, yielding a better specificity. This approach allows us to attach, for the first time, probabilities to found peaks, based on the signal strength of a peak across all sub-spectra.

References

[1] Pan Du, Warren A. Kibbe, and Simon M. Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22:2059–2065, 2006.

[2] Eva Lange, Clemens Gropl, Knut Reinert, Oliver Kohlbacher, and Andreas Hildebrandt. High-accuracy peak picking of proteomics data using wavelet techniques. *Pacific Symposium on Biocomputing*, 11:243–254, 2006.