Individual Abstract Info

Session: Bioinformatics **Code:** ThP32 **Time Slot/Poster Number:** 516

Automatic biomarker and compound identification using mass spectral data

Juergen H. Cox; Marc K. Flesch; Peter P. Haberl; Ruediger Heidenblut; Melanie Markmann; Jim O. Samuelsson;

Genedata, Martinsried, Germany

Introduction:

We present a set of novel methods which improve the performance of biomarker detection followed by subsequent compound identification. The input data are mass spectra (datapoints: [m/z, intensity] pairs) or chromatograms (datapoints: [m/z, retention time, intensity] triplets) pertaining to different samples or conditions (health, disease, treatments etc). The output is detected biomarkers and identified compounds. The general nature of these methods make them applicable to proteomics as well as metabolomics research. Furthermore, the methods are ideally well-suited for a streamlined and automated workflow. The methods and algorithms, which are designed for high throughput of very large data sets, have been implemented into available software.

Methods:

The data first go through a QA (quality assessment) step, where the m/z and/or the retention time dimensions are checked, calibrated and aligned. Baseline subtraction is also performed. The QA ensures that the subsequent comparisons between spectra or between chromatograms are performed on a common numerical grid and scale. The QA is followed by a peak extraction step where the abundance of every selected peak (area under curve) is fed into a data matrix. Multivariate analysis is then performed on the data matrix, resulting in a set of those peaks that discriminate the most between the samples. Finally, a spectral library or sequence database is searched for identification of the discriminating compounds (proteins, peptides, metabolites).

Abstract:

---

close window