9TH ANNUAL CONFERENCE OF THE METABOLOMICS SOCIETY

FULL ABSTRACTS

PLENARY SPEAKER ABSTRACTS

ORAL ABSTRACTS

POSTER ABSTRACTS

LATE POSTER ABSTRACTS

P9-49

THE USE OF MASS DEFECT FOR UNKNOWN BIOMARKERS RECOGNITION, CLASSIFICATION AND CHEMICAL FORMULA PREDICTION WITHIN LC-MS UNTARGETED EXPERIMENTS

Luca Narduzzi, Fulvio Mattivi

Edmund Mach foundation, Trento, Trentino, Italy

Typical LC-MS-based untargeted metabolomic experiments follow this workflow: 1) spectral data collection using high-resolution mass spectrometers, after separation of the analytes by UHPLC; 2) features alignment, extraction and putative grouping using ad-hoc algorithms (XCMS, METalign and many others); 3) putative biomarkers identification through statistical softwares; 4) biomarkers identification, confirmation and semi-quantification by injection of standards of known compounds; 5) *unknown* biomarkers annotation through assignment of a putative formula via MS-libraries search, MS/MS and MS^n data mining using softwares for mass spectral interpretation and successive manual curation. In particular, when dealing with plants extracts, the unknown compounds are >80% of the extracted signals. In the last decades many bio-informaticians improved the annotation of the *unknown* signals extracted from LC-MS untargeted experiments, but the chemical structure elucidation is still a complex and error-prone process. Data requires manual checking, identification and confirmation of the putative chemical formulas is time consuming and the MS/MS data are not always univoque or even available.

In this work, using the data coming from an untargeted metabolomic experiment on grapes, we employed a further step of systematic data visualisation in order to classify the *unknown* biomarkers signal using the *mass defect ratio* (MDR) and the *retention time* (RT) to create a 2D plot. The advantage of the plot is that it gives infomation about the chemical nature of the *unknown* compound; in facts, these parameters are affected from the chemical formula (both MDR and RT) and the chemical structure (RT). This plot allows a clear separation of the different *unknown* biomarkers. When the appropriate reference standards of the main classes are inserted, the MDR*RT plot gives useful information about the putative classes of the biomarkers, helping in excluding improbable formulas and structures. This MDR*RT plot, in combination with the latest rules for formula generation (seven golden rules, isotopic mass defect and others) allows to obtain a restricted pool of suggested formulas, ranking the correct one in the first positions and suggesting the chemical class of the compound. The joint use of online chemical databases and MS/MS data gives a satisfactory tentative identification of the experimental *unknown* biomarkers.