A workflow for data integration, analysis, and metabolite annotation for untargeted metabolomics

G. Frigerio[1,*], R. Mercadante[1], L. Campo[1], E. Polledri[1], L. Olgiati[1], P. Missineo[1], L. Boniardi[1], W.J. Nash[2], W.B. Dunn[2], S. Fustinoni[1]

[1] *EPIGET - Epidemiology, Epigenetics, and Toxicology Lab, Department of Clinical Sciences and Community Health, University of Milan and Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, via S. Barnaba, 8, 20122, Milan, Italy.*

[2] *School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.*

** Presenting author. E-mail: gianfranco.frigerio@unimi.it*

Metabolomics is the youngest of the "omics" disciplines and it is regarded as a promising approach to understand the metabolic changes that can occur in particular conditions and to identify new biomarkers. We present here a workflow for data integration, analysis, and metabolite annotation to be applied to untargeted metabolomic experiments.

Data acquired with LC-MS/MS, operating in data dependent mode, are processed using the R-packages IPO and XCMS to perform feature detection, retention time correction and alignment. The data-table obtained is elaborated and submitted to statistical analysis using the on-line software MetaboAnalyst. Multivariate analysis, in particular principal component and partial least squares discriminant analysis are performed for data visualization. Univariate analysis, in particular T-test for pairwise and ANOVA for multi-groups comparison, are performed to detect significant features among groups. The software BEAMS, developed by the University of Birmingham, is then implemented for grouping adducts and isotopes, and to perform a first annotation. Metabolite annotation is finally completed by comparing the fragmentation pattern obtained from each parent ion corresponding to a significant feature with data stored in on-line databases as Metlin, and with the help of the software MS-FINDER, which performs in-silico fragmentation.

We applied this workflow to an untargeted metabolomic experiment performed on 67 urine samples obtained from adult subjects with different smoking habits: non-smokers, electronic cigarette smokers, and traditional tobacco smokers. 117 features, out of 3613, were statistically different among groups. We estimated that they correspond to about 80 metabolites. We were able to putatively annotate compound classes of most of the significant metabolites (level 3 according to the "Proposed minimum reporting standards"; Sumner et al., 2007) and to putatively annotate some of them (level 2). Among them, the glucuronide conjugated of 3-hydroxycotinine supports the validity of the proposed approach.