BITS :: Call for Abstracts 2022 - Oral communication

Type	Oral communication
Session	Algorithms for Bioinformatics
Title	Metabolite identification in tandem MS untargeted metabolomics studies
All Authors	Alavikakhki Z(1,2,4), Drago D(2), Ulaszewska M(2), Bellini E(3), Mannella V(2), Morelli MJ(3), Andolfo A(2), Mosca E(4)

Affiliation

(1) Università degli Studi di Milano, Department of Biosciences, Milan
(2) ProMeFa, Proteomics and Metabolomics Facility, Center for Omics Sciences (COSR), IRCCS San Raffaele Scientific Institute, Milan
(3) Center for Omics Sciences (COSR), IRCCS San Raffaele Scientific Institute, Milan
(4) National Research Council, Institute of Biomedical Technologies, Segrate (Milan)

Motivation

Mass spectrometry (MS)-based metabolic profiling in circulating biofluids and tissues is a promising approach to identify biomarkers for disease prediction, progression and prognosis. Besides, metabolomics is a powerful methodology, which can be applied to personalized medicine for drug development or for biomarker discovery in public healthcare. However, the bottleneck in MS-based untargeted metabolic profiling lies in the availability of computational techniques able to process from beginning to end the very large amount of data in the post-acquisition steps. In particular, metabolite identification – the process of giving a known identity to the tens of thousands of features detected by untargeted MS profiling – is a challenging task. This is particularly true in tandem MS (MS/MS), which provides the MS/MS spectrum for each precursor ion characterized by a particular retention time (RT) and mass-to-charge ratio (m/z). Indeed, this step requires reference libraries obtained in experimental conditions that are influenced by a series of factors, like the technology type, vendor, analyzer type and chromatographic column. Besides, experimental errors complicate the matching between the metabolites of a novel dataset and those of a reference library. We present a pipeline that provides a complete analysis workflow for metabolic profiling, specifically addressing the challenge of metabolite identification in liquid chromatographic (LC)-MS/MS untargeted metabolomics studies.

Methods

The pipeline is implemented in R and is part of the R package “margheRita”. The metabolite identification is implemented in the following functions: “select_library”, which loads the appropriate reference, considering positive or negative polarity; “check_RT”, which quantifies the RT error; “check_mass”, which quantifies m/z ppm (part per million) error; “check_peaks”, which quantifies the m/z ppm error among all-pairs of fragments in MS/MS spectra; “metabolite_correlation”, which quantifies the correlation between metabolites; “visualize_spectra”, which visualizes two spectra; “metabolite_identification”, a wrapper that executes the full pipeline. The code is implemented in parallel. The metabolite library was created with analytical standards from Sigma Aldrich “Mass Spectrometry Metabolite Library” on 4 different chromatographic column types: RP-C18, HILIC, RP-C8 and pZIC-HILIC Zwiterrionic. Analytical standards were injected on 5 different chromatographic gradients, which enable the screening of different metabolite classes: lipids, semi-polar, highly-polar and zwitterionic. The original sample set pzic-HILIC was obtained processing urine samples from kinetic experiment where 3 subjects (3 technical replicates each) consumed 3 different meals. Kinetics data were collected over 8 time points. Raw MS/MS data were extracted (peak picking, peak alignment and MS/MS SWATH deconvolution) through MS Dial software.

Results

The annotation of features is performed against the in-house reference library described above, separately for positive and negative ionization modes acquired on our instrument. This library provides “Level 1” annotation, which includes RT, m/z accurate mass for molecular ion, and MS/MS fragmentation, along with a series of metabolite descriptors, such as Common Name, CAS number, ChEBI, PubChemID. These descriptors facilitate downstream analyses, like pathway analysis. The annotation process consists of two main steps: firstly, the given m/z features are screened by matching their RT and mass accuracy with those of the metabolites in the library. Secondly, the MS/MS spectra of features that passed positively the first screening are compared with those of the library, considering one or more of the most intense peaks in each spectrum. This two-steps procedure significantly saves the computational effort. The process of metabolite annotation can be controlled acting on a series of parameters (e.g., error thresholds, selection thresholds, number of peaks to analyze) that allow the analyst to explore and optimize the process according to the particular conditions of the study. Our pipeline also provides two approaches to improve the identification of remaining unknown compounds. By assessing the similarity of MS/MS spectra, we indicate features exhibiting similar behavior to the reference metabolite, in terms of mass accuracy of fragment ions and their intensity without considering the RT. Finally, by studying the correlation of metabolite levels across samples, we infer a functional similarity (e.g. same pathway) between an unknown metabolite and metabolites with a known identity. We illustrate our pipeline using urine samples generated on purpose for this study, in order to show its applicability to identify metabolites at level 1, and the explorative function for unknown compounds.

Info

-	
filename	-
Figure	
-	
Availability	-

Corresponding Author	
Name, Surname	Ettore, Mosca
Email	ettore.mosca@itb.cnr.it
Submitted on	11.05.2022