Data in Brief 3 (2015) 155–164

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data in support of enhancing metabolomics research through data mining

Ibon Martínez-Arranz [a], Rebeca Mayo [a], Miriam Pérez-Cormenzana [a], Itziar Mincholé [a], Lorena Salazar [b], Cristina Alonso [a], José M. Mato [c,*]

[a] OWL, Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain
[b] Osarten kooperatiba elkartea, Mondragón, Guipúzcoa, Spain
[c] CIC bioGUNE, CIBERehd, Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain

ARTICLE INFO

Article history:
Received 5 February 2015
Accepted 9 February 2015
Available online 27 February 2015

ABSTRACT

Metabolomics research has evolved considerably, particularly during the last decade. Over the course of this evolution, the interest in this 'omic' discipline is now more evident than ever. However, the future of metabolomics will depend on its capability to find biomarkers. For that reason, data mining constitutes a challenging task in metabolomics workflow. This work has been designed in support of the research article entitled "Enhancing metabolomics research through data mining", which proposed a methodological data handling guideline. An aging research in healthy population was used as a guiding thread to illustrate this process. Here we provide a further interpretation of the obtained statistical results. We also focused on the importance of graphical visualization tools as a clue to understand the most common univariate and multivariate data analyses applied in metabolomics.

DOI of original article: http://dx.doi.org/10.1016/j.jprot.2015.01.019
* Corresponding author.
E-mail address: director@cicbiogune.es (J.M. Mato).

http://dx.doi.org/10.1016/j.dib.2015.02.008

Specifications Table

Subject area	*Chemistry/Biology*
More specific subject area	*Human metabolomics.*
Type of data	*Table, R code files, graph, figure.*
How data was acquired	*Mass spectrometry, clinical laboratory.*
Data format	*Comma-separated values (*.csv) tables.*
Experimental factors	*Serum samples from healthy male and female, collected under fasting conditions.*
Experimental features	Methanol and chloroform/methanol serum extracts were analyzed with three separate ultra-performance liquid chromatography-mass spectrometry based platforms.
Data source location	*Basque Country, Spain.*
Data accessibility	*Data are available here and via a web application (http://rstudio.owlmetabolomics.com:8031/AgingAnalysis/)*

Value of the data

- Metabolites related to aging in healthy population are highlighted as a result of two different post-acquisition approaches, considering age as a categorical and a continuous variable.
- R functions are provided for different statistical test, including graphical visualization tools.
- Data are presented through a web application. This is expected to help with the visualization and interpretation of univariate and multivariate data analyses.

1. Data

Serum samples and anthropometric data from healthy male and female volunteers included in this study were provided by the Basque Biobank for Research-OEHUN (http://www.biobancovasco.org/) and were processed with appropriate approval of the Ethics Committee. Samples were analyzed in a COBAS 6000 (Roche Diagnostics GmbH, Germany) and hematological parameters in a GEN-S (Beckman COULTER Inc., USA) at OSARTEN K.E. laboratory.

Metabolomics profiling data acquired by ultra-performance liquid chromatography coupled to mass spectrometry (UPLC-MS) were pre-processed using the TargetLynx application manager for MassLynx 4.1 (Waters Corp., Milford, MA). The peak-picking process included 466 metabolic features, identified prior to the analysis.

Then, all calculations were performed using R v.3.1.1 (R Development Core Team, 2011; http://cran.r-project.org) [1].

2. Experimental design, materials and methods

In metabolic profiling, there is no single platform or method to analyze the entire metabolome of a biological sample, mainly due to the wide concentration range of the metabolites coupled to their extensive chemical diversity [2,3]. The current study used multiple UPLC-MS platforms, which were optimized for extensive coverage of the serum metabolome. Metabolite extraction was accomplished by fractionating the samples into pools of species with similar physicochemical properties, using appropriate combinations of organic solvents [4]. Then, three separate UPLC-MS based platforms were used. Briefly, UPLC-single quadrupole-MS amino acid analysis system was combined with two separate UPLC-time-of-flight-MS based platforms analyzing methanol and chloroform/methanol extracts. Identified ion features in the methanol extract platform included non-esterified fatty acids,