A MULTI-OMICS BIOINFORMATICS WORKFLOW FOR THE INTEGRATION AND INTERPRETATION OF TRANSCRIPTOMICS AND METABOLOMICS DATA

Nafsika Papaioannou[1,2,5], Thanasis Papageorgiou[1,2], Dayna R Schultz[1,2], Ioannis Dallas[1,2], Ilias Frydas [1,2], Catherine Gabriel[1,2,5], Spyros Karakitsios[1,2,5] , Dimosthenis Sarigiannis[1,2,3,4]

[1] Environmental Engineering Laboratory, Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece.
[2] HERACLES Research Center on the Exposome and Health, Center for Interdisciplinary Research and Innovation, Balkan Center, Bldg. B, 10th km Thessaloniki-Thermi Road, Thessaloniki, Greece.
[3] Environmental Health Engineering, Institute of Advanced Study, Palazzo del Broletto- Piazza Della Vittoria 15, 27100, Pavia, Italy.
[4] National Hellenic Research Foundation, 48 Vasileos Constantinou Ave., Athens, Greece
[5] ENVE.X, Thessaloniki, Greece https://enve-x.com/
(*denis@eng.auth.gr)

ABSTRACT

Advancements in high-throughput technologies, characterizing various biological "omes" in samples, emphasize the need for integrative algorithms facilitating data unification and knowledge discovery. This article presents a robust workflow developed in the R/Bioconductor architecture for integrating and interpreting transcriptomics and metabolomics data, validated with a dataset from pancreatic EndoC-βh1 cells exposed to bisphenols (BPA, BPF, and BPS). The workflow encompasses sample preparation, data acquisition protocols, and data analysis processes for individual omics data, followed by a detailed workflow for multi-omics analysis.

For transcriptomic analysis, the One-Color Microarray-Based Gene Expression Analysis Protocol was followed, ensuring optimal microarray results. Untargeted metabolomics analysis was performed using an LC-QTOF-HRMS system using Reversed Phase (RP) and Hydrophilic Interaction (HILIC) Liquid Chromatography in positive and negative ionization modes. Transcriptomics raw data were exported as .txt files and imported into R. In R, data were analyzed using the limma package with built-in analyses specific for One-Color Agilent microarray data. The untargeted metabolomics data were processed using Bioconductor Software Packages (e.g XCMS, IPO, and CAMERA). The Kruskal-Wallis test was applied after multi-filtering, normalization, scaling, log transformation, and batch effects correction, to detect differentially expressed features (DEF). DEF were annotated using online and in-house compound databases.

The multi-omics workflow is based on MetaboAnalystR [1] and MixOmics [2] Bioconductor Software Packages, incorporating multivariate analysis, and joint pathway analysis. Integration of bioinformatics approaches, including network analysis, KEGG Mapper, Pathvisio, and Reactome, was employed for additional mapping. Multi-omics pathway analysis of 3D HepaRG cells exposed to pollutants identified 142 dysregulated pathways related to oxidative stress and lipid metabolism.

Overall, the multi-omics analysis revealed perturbations in various pathways involved in amino acid and lipid metabolism that have been linked with downstream effects such as obesity, diabetes, and NAFLD.

In conclusion, the presented multi-omics workflow in R provides a valuable tool for identifying potential biomarkers associated with the toxicity mechanisms of investigated pollutants. The integration and interpretation of omics data contributes to a mechanistic understanding of underlying processes, crucial for translating findings into risk assessments.

KEYWORDS: Systems biology, Multi-omics, bioinformatics, *in vitro*, bisphenols

REFERENCES

[1] Pang Z, Chong J, Li S, Xia J. (2020) MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. Metabolites. 10 (5), 32392884

[2] Rohart F, Gautier B, Singh A, Lê Cao K-A. (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Computational Biology, 13, 1005752