A COMPARISON OF THREE METHODS FOR MARKERS SELECTION IN UNTARGETED LIPIDOMICS

Polanco Espino F.J.[1], Demicheva E.I.[1, 2], Ushenin K.S.[1, 2], Solovyova O.E.[1, 2]
[1)] Ural Federal University named after the First President of Russia B.N. Yeltsin
[2)] Institute of Immunology and Physiology of the Ural Branch of the Russian Academy of Sciences
E-mail: fpolancoespino@gmail.com

In the present paper is analyzed different algorithms for the marker's selection of mass-spectrometry data in lipidomics. The goal of the investigation is to highlight the shortcomings and advantages of the processes and determine the most suitable.

In the present paper is analyzed different algorithms for the marker's selection of mass-spectrometry data in lipidomics. The goal of the investigation is to highlight the shortcomings and advantages of the processes and determine the most suitable.

High-performance liquid chromatography-Mass Spectrometry (HPLC-MS) is a powerful analytical technique used for identification of structure and chemical properties of different molecules[3]. The process of this data is divided into 4 steps, pre-processing, filtration, lipid identification, and statistical analysis. For this study a path composed of the XCML and the Lipidfinder algorithms for preprocessing, filtration and lipid identification step was selected.

The present paper is aimed to analyze the use of 3 different paths for markers selection in mass spectrometry data in lipidomics. In the present work we selected 3 different paths for the statistical analysis and compared them to select the most suitable. The first two paths are composed by different steps, normalization, scaling, correlation analysis and volcano plot. One of them uses vector length normalization[1] and the second one quantile normalization[2]. The third path uses decision trees. In the three cases the goal was to determine the statistical significance of the results and differentiate the substances with a greater difference between both groups.

The data used was of 10 mice, 5 of them were taken with immunodeficiency state, whereby other 5 animals were healthy. It is characterized by a large size and a big dispersion in the intensity ranges of each of the identified substances, which makes it difficult to identify the most significant of them in the statistical analysis.

The research funding from the Ministry of Science and Higher Education of the Russian Federation (Ural Federal University Program of Development within the Priority-2030 Program) is gratefully acknowledged.

1. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal. Chem. 78, 779–787 (2006)
2. Alvarez-Jarreta, J. et al. LipidFinder 2.0: advanced informatics pipeline for lipidomics discovery applications. Bioinformatics 37, 1478–1479 (2021)

3. Mukherjee, P. K. Chapter 11 - LC–MS: A Rapid Technique for Understanding the Plant Metabolite Analysis. in Quality Control and Evaluation of Herbal Drugs (ed. Mukherjee, P. K.) 459–479 (Elsevier, 2019). doi:10.1016/B978-0-12-813374-3.00011-9
4. Fahy, E. et al. LipidFinder on LIPID MAPS: peak filtering, MS searching and statistical analysis for lipidomics. Bioinformatics 35, 685–687 (2019)
5. Xi, B., Gu, H., Baniasadi, H. & Raftery, D. Statistical Analysis and Modeling of Mass Spectrometry-Based Metabolomics Data. in Mass Spectrometry in Metabolomics: Methods and Protocols (ed. Raftery, D.) 333–353 (Springer New York, 2014). doi:10.1007/978-1-4939-1258-2_22