Contribution ID: **10** Type: **not specified**

Extracting biological meaning from lipidomics data through biostatistics, machine learning and pathway analysis

Wednesday 22 November 2023 11:55 (30 minutes)

In recent years, due to significant advancements in mass spectrometry, lipidomics has emerged as a fast growing scientific field that provides deep insights into complex changes in lipidome of human cells and tissues. Human blood is a self-regenerating lipid-rich biological fluid, a tissue that is easily collected in hospitals. Blood/plasma is rich in lipids and related metabolites, and its lipid composition (LIPIDOME) reflects diverse aspects of lipid metabolism and may give us insight into general human physiology in health and disease. Plasma lipidome is a tightly regulated and precisely defined constellation of lipid molecules and disturbances in the plasma lipidome occur in many diseases such as cardiovascular diseases and cancer, but also in conditions that are not directly linked to lipid metabolism. Hence, plasma lipidomics is an emerging tool in an array of clinical diagnostics. The most sought-after goals in the lipidomics community are to identify disease biomarkers, monitor a clinical treatment or confirm a biological hypothesis on a causality between a disease onset/progression and lipid profile. A recommended plasma lipidomics workflow consists of: preanalytics, analytics and post analytics, including study design, research hypotheses, sample collection, demographics data collection, lipid extraction, quality control, liquid chromatography and mass spectrometry, raw data processing, lipid annotation/identification, normalization, lipid quantitation, databases, data sharing and data analysis. The topic of this presentation is data analysis so I will put an accent to it. Lipidomics community nowadays uses dozens of biostatistical tools, artificial intelligence (AI), machine learning (ML) algorithms and chemometrics for data analysis, depending on size of datasets, data structure and the expertise of data scientists involved. Smart analysis of lipidomic data (provided these are of good quality) is of paramount importance for identifying potential biomarkers and understanding disease mechanisms. To accomplish this, they use, besides classical statistics-between groups comparisons, principal component analysis (PCA), orthogonal projections to latent structures discriminant analysis (OPLS-DA), partial least square analysis (PLS) and correlation analysis, all accompanied by various data pattern recognition and data visualization tools. Less often used are ML algorithms such as random forests (RF) and support vector machines (SVM), with only few applying deep learning and neural networks. Finally, to extract biological knowledge, i.e. metabolic pathways affected by a disease or applied treatment, data are subjected to different enrichment analyses, network analyses and pathway analysis. Life scientists working outside clinical setting, such as those investigating neurodegenerative changes in brain will also benefit from applying lipidomics to their model systems.

Primary author: Dr MASNIKOSA, Romana (Vinča Nuclear Research Institute)

Presenter: Dr MASNIKOSA, Romana (Vinča Nuclear Research Institute)