MASS SPECTROMETRY-BASED METABOLOMICS IS A CENTREPIECE IN THE ANALYSIS OF BIOLOGICAL MATRICES AND THE DISCOVERY OF NEW BIOMARKER METABOLITES WITH APPLICATIONS IN HUMAN DISEASE DIAGNOSTICS

A basis for metabolomics

Biomarkers that can accurately detect early stage disease and differentiate it from later stages are needed in all therapeutic areas. While extensive progress has been made in the fields of genomics and proteomics, additional evidence of the biological end points of human diseases is highly desired for disease diagnosis and prognosis, as well as therapeutic development. The detection of metabolites involved in human diseases can be used to develop new biomarkers, using cells, tissues, organs or biological fluids. Among the analytical techniques that can be employed in metabolomics applications, mass spectrometry (MS) in combination with separation techniques such as liquid chromatography (LC) enables the quantitative detection of multiple small molecule metabolites, providing an efficient method for monitoring changes in concentrations and fluxes of specific endogenous metabolites involved in key disease-related or other specific cellular pathways.

Mass spectrometry-based approaches for metabolomics

In the field of metabolomics, the efficacy of MS-based approaches has long been limited by sensitivity and metabolite identification ability. However, low resolution instruments such as triple quadrupoles are still widely used for targeted approaches in which they enable sensitive detection and quantification of tens of hundreds of specific metabolites. The recent development of affordable high resolution mass spectrometry (HRMS) has overcome the former limitations of the sensitivity and identification ability. It is thus possible to detect metabolites which have the same nominal mass but distinct exact masses in complex biological matrices, and also to propose their chemical composition, facilitating database searches for identification purposes.

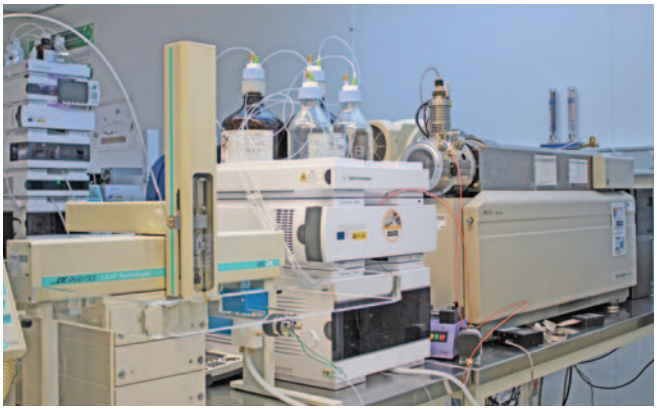
MEDINA's mission is the discovery of biologically active molecules, drug candidates for unmet medical needs and products of high biotechnological value for the industry

This fact has enabled HRMS to become one of the most widely used tools for untargeted biomarker discovery approaches. Currently, there are two main classes of high resolution mass spectrometers: time-of-flight (TOF) and Orbitrap-based instruments. Whereas Orbitrap instruments still provide the best results in terms of accuracy and mass resolving power, TOF instruments provide high resolving power, independently of the acquisition rate, which could be considered a crucial feature, especially in the case of fast chromatographic separations. Irrespective of these differential characteristics, both systems have been described as equally efficient for untargeted metabolomics.

The LC/HRMS-based metabolomics workflow

The usual metabolomics workflow can be divided into several stages that involve analytical optimisation, sample preparation for metabolite extraction, data acquisition, data processing, data annotation and statistical analysis. Setting up adequate analytical methods is a key initial step to minimising the effect of artefactual correlations of analytical origin. Following their acquisition, data are preprocessed to filtrate noise signals and artefacts. As a result, a data matrix is obtained containing the m/z value and retention time of each feature and their abundance in the sample. Thanks to the implementation of normalisation algorithms which enable correcting intra and inter-batch effects, LC/MS-based approaches have become compatible with the analysis of a large series of samples typically required in biomarkers discovery.

In metabolomic studies focused on deciphering biological processes, data analysis is usually performed using multivariate statistical methods such as principal component analysis (PCA) or partial least squares discriminant analysis (PLS-DA). These dimension reduction methods summarise and transform hundreds to thousands of metabolite features into a few key components that capture the maximal variance



or discriminatory covariance in the data. In most cases, these statistical analyses are not ideal for clinical biomarker discovery, which requires slightly different analysis, evaluation and validation procedures. The receiver operator characteristic (ROC) curves analysis is usually accepted as the standard procedure for assessing the performance of medical diagnostic tests. An ROC curve depicts sensitivity *versus* specificity, and their area under the curve (AUC) can generally be interpreted.

The utility of a biomarker can be assessed based on its AUC, which might range from 1.0 (excellent) to 0.5 (fail), in order to select and combine them into a single multivariate model that provides the required high levels of discrimination and confidence. Lastly, the best way to assess the potential clinical utility of a candidate biomarker is a validation experiment which consists of repeating the experiment on independent samples drawn from the same target population.

Biomarker selection

The feature selection stage involves optimising two criteria: 1) the biomarker utility – AUC etc.; and 2) the number of metabolites used in the predictive model. There are two main strategies for this multi-objective optimisation in metabolomics. Feature filtering is the most straightforward and commonly applied method of feature selection in metabolomics projects and only requires ranking the variables used in the model according to their discrimination ability, and then reprocessing the model using the top discriminating variables. The optimal number of variables is then subjectively selected to yield an appropriate ROC curve.

The strength of this approach lies in its simplicity, which does not demand an extensive understanding of modelling algorithms. However, it might often yield a model with over-fitted performance, especially for projection methods, where there is an inherent dimensionality reduction. The wrapper method is the other main approach for feature selection in metabolomics. Roughly, it consists of adding variables to the model, one at a time, until there is no significant improvement to it. Different versions of this approach have been developed based on computationally intense selection methods to exhaustively search all combinations of available variables until the best model is settled.

Metabolite identification

The mass measurements provided by HRMS enable metabolite identification using spectral databases including accurate mass, and also automatic matching of experimental masses with those contained in chemical and metabolomic databases such as the Human Metabolome Database (HMDB), METLIN, KEGG or PubChem. However, accurate mass data is often insufficient for tentative metabolite identification, and collision-induced dissociation experiments to obtain fragmentation spectra are required in order to characterise unknown metabolites.

Perspectives

Metabolomics has become a major component of system biology, and MS-based applications focused on early disease detection have experienced an exponential growth. In particular, metabolomics studies focused on the identification of metabolites associated with diseases such as cancer, diabetes, inborn errors of metabolism and cardiovascular diseases are providing promising results. However, validation studies are still in progress to confirm the extensive clinical use of these candidate biomarkers.

MEDINA is a non-profit research organisation established in the Health Sciences Technology Park, Granada (Spain), as a private public partnership between the government of Andalucía, the University of Granada and Merck Sharp and Dohme de España S.A. Leveraging our unique strengths in high precision, high throughput analytics and bioanalytics, we collaborate with clinical and academic partners from several institutions in translational research programmes for biomarker discovery and metabolomic profiling associated with different diseases (diabetes, cardiovascular diseases, cancer, anti-infectives) and with applications in the area of microbial natural products metabolomics. Our LC-MS metabolomics platform provides state-of-the-art capabilities in terms of sample preparation, instrumentation (LC-QTOF and LC-triple quadrupole configurations for metabolite profiling and targeted metabolite profiling, respectively), innovative model building and access to proprietary spectral databases for metabolite identification.

Jose Pérez del Palacio
jose.perezdelpalacio@medinaandalucia.es

Francisca Vicente
francisca.vicente@medinaandalucia.es

Olga Genilloud
olga.genilloud@medinaandalucia.es
Fundación MEDINA

+34 958 993 965

www.medinadiscovery.com