A GENERAL STATISTICAL FRAMEWORK FOR THE INTEGRATION AND ONTOLOGICAL ANALYSIS OF QUANTITATIVE PROTEOMICS EXPERIMENTS BY STABLE ISOTOPE LABELLING

M.Trevisán-Herraz[(1)], E.Bonzón-Kulichenko[(1)], P.Navarro[(2)], P.Martínez-Acedo[(1)], E.Núñez[(1)], D.Pérez-Hernández[(1)], I.Jorge[(1)], R.Mesa[(1)], E.Calvo[(3)], M.Carrascal[(4)], M.L.Hernáez[(5)], F.García[(6)], K.Ashman[(6)], J.Abián[(4)], C. Gil[(5)], J.M.Redondo[(3)] and J.Vázquez[(1)].

[(1)]Centro de Biología Molecular Severo Ochoa-CSIC. [(2)] Institute of Molecular Systems Biology, Zurich, Suiza. [(3)]Centro Nacional de Investigaciones Cardiovasculares. [(4)]Instituto de Investigaciones Biomédicas de Barcelona. [(5)]Universidad Complutense de Madrid. [(6)]Centro Nacional de Investigaciones Oncológicas.

Current MS-based proteomics techniques can identify and quantify thousands of proteins. Nevertheless, extracting a reliable list of proteins that significantly change their expression, determining the accuracy of this information, and interpreting the biological relevance of this plethora of data is still an open problem.

In a previous work, we presented a random effects hierarchical model for the analysis of protein expression changes in ^{18}O-labelling/LIT experiments. This statistical model for the null hypothesis splits the total variance into three sources (spectrum, peptide and protein) and has been tested on several proteomes of different nature. The model has been implemented into QuiXoT, a software platform developed in our laboratory. Furthermore, in a collaborative, large-scale project using H_2O_2-treated *Saccharomyces cerevisiae* as a model system, we demonstrated this model is also valid for other SIL techniques, including SILAC, iTRAQ and different MS instruments.

Here we have further developed the model and demonstrated that it also allows a coherent integration of quantitative protein information and variance from multiple experiments, irrespective of labelling or MS technique used, allowing not only the detection of significant protein expression changes but also a full control of inter-experiment variability. Moreover, we also demonstrated that the extended model allows a threshold-free analysis of all the data, enabling a high-power characterisation of quantitative behaviour at the level of ontological categories.The extended model offers for the first time a coherent and universal statistical framework for the analysis of quantitative proteomics data obtained by SIL approaches in any MS instrument, and a straightforward integration with Systems Biology tools, facilitating the interpretation of results in terms of biological significance.