

estricto control sobre las posibles fuentes de error del experimento.

El desarrollo de un modelo estadístico de validez universal y de una plataforma informática común permite por primera vez el análisis sistemático y la detección de cambios de expresión significativos de forma semiautomática en cualquier experimento de expresión diferencial a escala masiva mediante técnicas de marcaje isotópico estable.

## References

- [1] Jorge I, Navarro P, Martinez-Acedo P, Nunez E, Serrano H, Alfranca A, et al. Statistical model to analyze quantitative proteomics data obtained by 18O/16O labeling and linear ion trap mass spectrometry: application to the study of vascular endothelial growth factor-induced angiogenesis in endothelial cells *Mol Cell Proteomics* 2009;8:1130-49.
- [2] Bonzón-Kulichenko E, Pérez-Hernández D, Martínez-Acedo P, Núñez E, Navarro P, Trevisan-Herraz M, et al. A robust method for quantitative high-throughput analysis of proteomes by 18O labeling *Mol Cell Proteomics* 2009;En revisión.
- [3] Ramos-Fernández A, López-Ferrer D y Vázquez J. Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency *Mol Cell Proteomics* 2007;6:1274-86.

## Label-free Quantitative Approaches in CSF Biomarker Discovery

*Alex Campos<sup>1</sup>, Jacques Borg<sup>2</sup>, Claudio Diema<sup>3</sup>, Marta Vilaseca<sup>3</sup>, Eliande Oliveira<sup>1</sup>*

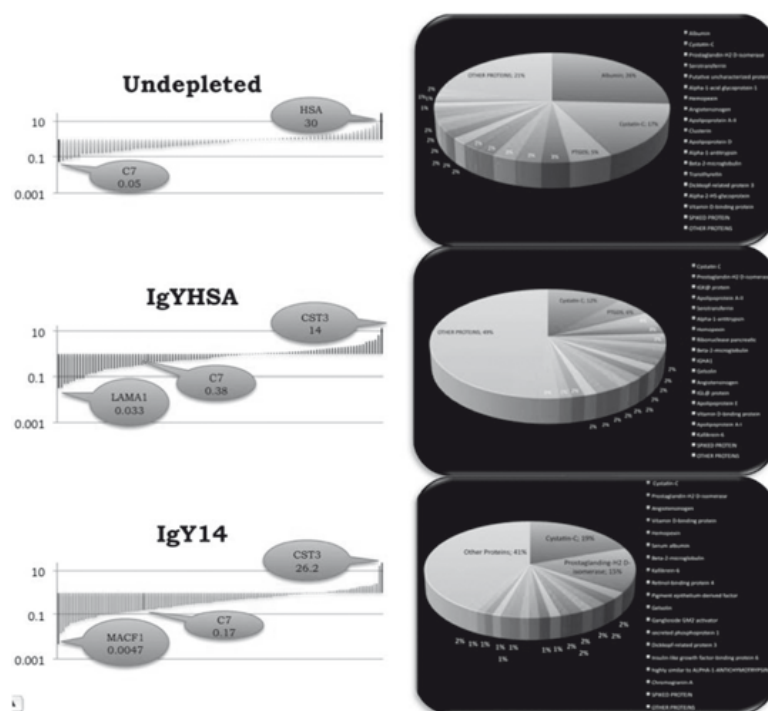
<sup>1</sup>Proteomics Platform, Barcelona Science Park, Barcelona, Spain. <sup>2</sup>Neurochemistry Lab, Faculty of Medicine, Saint Etienne, France. <sup>3</sup>Mass Spectrometry Core Facility, Institute for Research in Biomedicine, Barcelona, Spain

Mass Spectrometry (MS) has emerged as one of the most promising core technologies for discovery of diagnostic, prognostic and therapeutic protein biomarkers. Despite its rapid technological developments, MS-based assays have also raised major concerns regarding its reproducibility, sensibility and throughput, and hence its potential application in the clinical and regulatory settings. To address these concerns, we explore relevant quality assurance benchmarks for MS-based quantitative CSF biomarker discovery. We argue that the road to a reproducible and sufficiently robust biomarker development technology entails the understanding of the limits of each process, including sample handling, instrument setup, and bioinformatics tools.

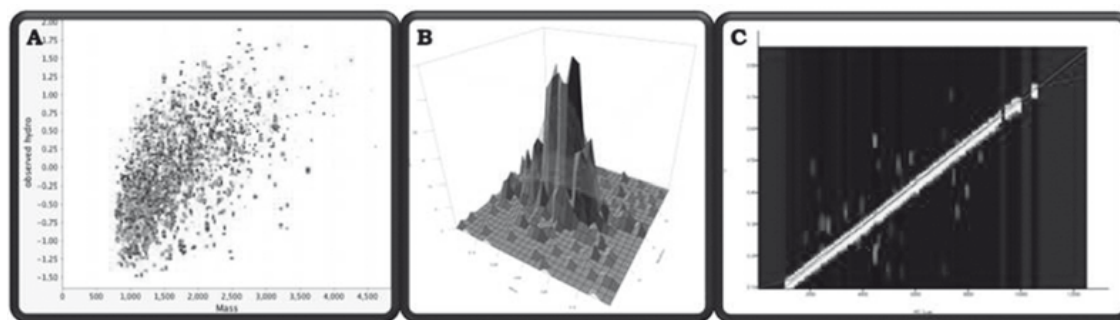
Despite the increasing clinical interest in biofluids as source for biomarker discovery, its distinctive nature imparts significant challenges for current proteomics technologies. The immense dynamic range of human plasma proteins, which spans 10 to 12 orders of magnitude, poses a major limiting factor for in-depth proteomics profiling. To overcome

this problem, enrichment techniques and orthogonal fractionation strategies are routinely applied in proteomics studies prior to MS analysis. Biomarkers of “interest” could be enriched with the selection of a biological material that is in close proximity with the disease source. In this regard, the cerebrospinal fluid (CSF) is a treasure trove for biomarker discovery in neurological diseases. Because of its direct contact with the brain’s extracellular space, CSF is a valuable reporter of processes that occur in the central nervous system. Although CSF shares some of the protein content of blood, the bona fide, locally produced brain’s proteins are found in higher concentration in the CSF compared to blood [1].

Platforms combining removal of high abundance proteins using multi-component immunodepletion systems (IDS) followed by multidimensional protein/peptide fractionation are currently the mainstay of unbiased proteomic biomarker discovery. At the present, a number of commercial IDS are available for highly selective removal of most abundant human plasma proteins. To the best



**Figure 1.** Distribution of the APEX values (left side) for the CSF proteome following depletion with IgY-HSA or IgY-14, and no depletion (y-axis on log scale). Pie charts (right side) depict the CSF protein content in context of their abundance. (HSA, human serum albumin; C7, complement 7; CST3, Cystatin C; LAMA1, Laminin- $\alpha$ -1 and MACF1, Microtubulin-actin)



**Figure 2.** AMT quality control metrics. (A) 2D display of the CSF AMT database. (B) 3D perspective plot of the peptide matches between IgY14-depleted sample and the AMT database (axis: mass and retention time error). (C) Heatmap of similarity score between LC-MS features of IgY14-depleted sample and AMT database. Best alignment is found using a dynamic programming algorithm that determines the transformation function with maximum likelihood.

of our knowledge, all commercial IDS have been devised to deplete plasma (or serum) samples and its performance in other biofluids has not been evaluated yet. To address this problem, we carried out CSF depletion (in triplicate) using either Sigma's Seppro systems designed to immunodeplete HSA (IgY-HSA) or the 14 most abundant proteins (IgY-14). Label-free computationally modified spectral counting method for Absolute Protein EXpression (APEX) measurements was used to assess the deple-

ted and undepleted samples in the context of protein content stoichiometry and achieved dynamic range (Figure 1). Our results show that IDS originally devised to deplete blood's most abundant proteins are deficient in deplete the most abundant CSF proteins. We therefore propose a list of target protein (most abundant) to be depleted in CSF for downstream MS profiling (Figure 1 – top, pie chart).

The advent of a new generation of robust MS instruments with very high resolution and mass ac-

curacy has inspired the development of new quantitative approaches [2]. In particular, label free hybrid identity/pattern-based approaches have gained popularity in the biomarker development field [3]. Such approaches use pattern recognition algorithms to *ex post facto* assign the identity of LC-MS peaks against databases of peptide sequence, mass, and retention time built from multiple experiments. Here we employ the accurate mass and time (AMT) strategy to identify and quantify peptides/proteins from unfractionated CSF samples using msInspect platform [4]. Our analysis workflow extends and combines existing open-source platforms for LC-MS/MS (Trans Proteomic Pipeline) and LC-MS (msInspect) data analysis. A peptide accurate mass and LC elution time AMT database was initially generated using MS/MS following extensive multidimensional LC separations to provide the basis for subsequent peptide identifications. Our CSF AMT database contains >2,000 entries from a total number of ~7,500 identified peptides, which translates into 558 confi-

dent CSF protein identifications (Figure 2). In conclusion, hybrid identity/pattern-based approaches are amenable for high throughput quantitative proteome analyses in biomarker discovery pipelines.

## References

- [1] Zougman, A et al. Integrated Analysis of the Cerebrospinal Fluid Peptidome and Proteome. *J Proteome Res* 2007; 7: 386-399.
- [2] Mueller, LK et al. An Assessment of Software Solutions for the Analysis of Mass Spectrometry Based Quantitative Proteomics Data. *J Proteome Res* 2008; 7: 51-61.
- [3] Jaffe, J et al. PEPPeR, a Platform for Experimental Proteomic Pattern Recognition. *Mol Cell Proteomics* 2006; 5:1927-1941.
- [4] May, D et al. A Platform for Accurate Mass and Time Analyses of Mass Spectrometry Data. *J Proteome Res* 2007; 6: 2685-2694.

## A comparison of quantitative proteomics methodologies on a differential experiment on test samples

Joan Josep Bech-Serra, Núria Colomé, Marta Monge, Francesc Canals

Laboratori de Proteòmica, Programa de Recerca en Oncologia Mèdica. Institut de Recerca. Hospital Vall d'Hebron, Barcelona

The field of proteomics includes a wide number of technologies aimed to the analysis of large number of proteins, representing ideally the entire proteome, in the same experiment. In its early stages, mass spectrometry-based proteomics was successfully used to the qualitative characterization of complex mixtures of proteins, leaving the quantitative aspects in a secondary role, essentially because of the lack of suitable technologies for quantitative analysis of such complex mixtures. However, in the last years, a wide number of strategies have been developed in order to give to the proteomics field robust tools to obtain reliable quantitative data. Those strategies include both 2D-gel based and non-based methodologies. In the first group, 2D-DIGE methodology is based on the labeling with different fluorochromes, enabling the simultaneous separation of different samples in the same 2D-gel, which allows to over-

come reproducibility problems inherent to the 2D procedure and internal standardization of spot volume measurements, providing a robust quantitative comparison technique. Alternatively, a number of gel-free quantitative techniques can be used. Some of them are based on non-isobaric isotope labeling, such as ICAT, ICPL or SILAC labeling methodologies, the quantitative data being obtained from the MS data, by integrating extracted ion chromatograms of the heavy-light peptide pair masses. A second group which includes iTRAQ or TMT reagents is based on isobaric labeling that introduce different reporter fragments in the derivatized peptides. Quantitative data is in this case obtained at the MS-MS level, from the relative intensities of the reporter fragment ions. Finally, a third group of label-free methods uses different approaches for quantitative comparison of LC-MS data.