

**Figure 2.** Graphical representation of identified proteins. A. Number of proteins identified in ECOLI\_1 and ECOLI\_2 per number of replicates. B. Venn diagram of overlapping proteins identified in 3 or more replicates in ECOLI\_1 and ECOLI\_2.

This work was supported by grants from the Basque Government Department of Industry (Etor-tek and Saiotek to J.M.A.). Proteomics Core Facility-SGIker is supported by UPV/EHU, MICINN, GV/EJ, ESF and ProteoRed agencies.

[1] Silva JC, Denny R, Dorschel CA, Gorenstein MV, Kass IJ, Li GZ, et al. Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* 2005;77:2187-200.

[2] Li GZ, Vissers JP, Silva JC, Golick D, Gorenstein MV and Geromanos SJ. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* 2009;9:1696-719.

[3] Silva JC, Gorenstein MV, Li GZ, Vissers JP and Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 2006;5: 144-56.

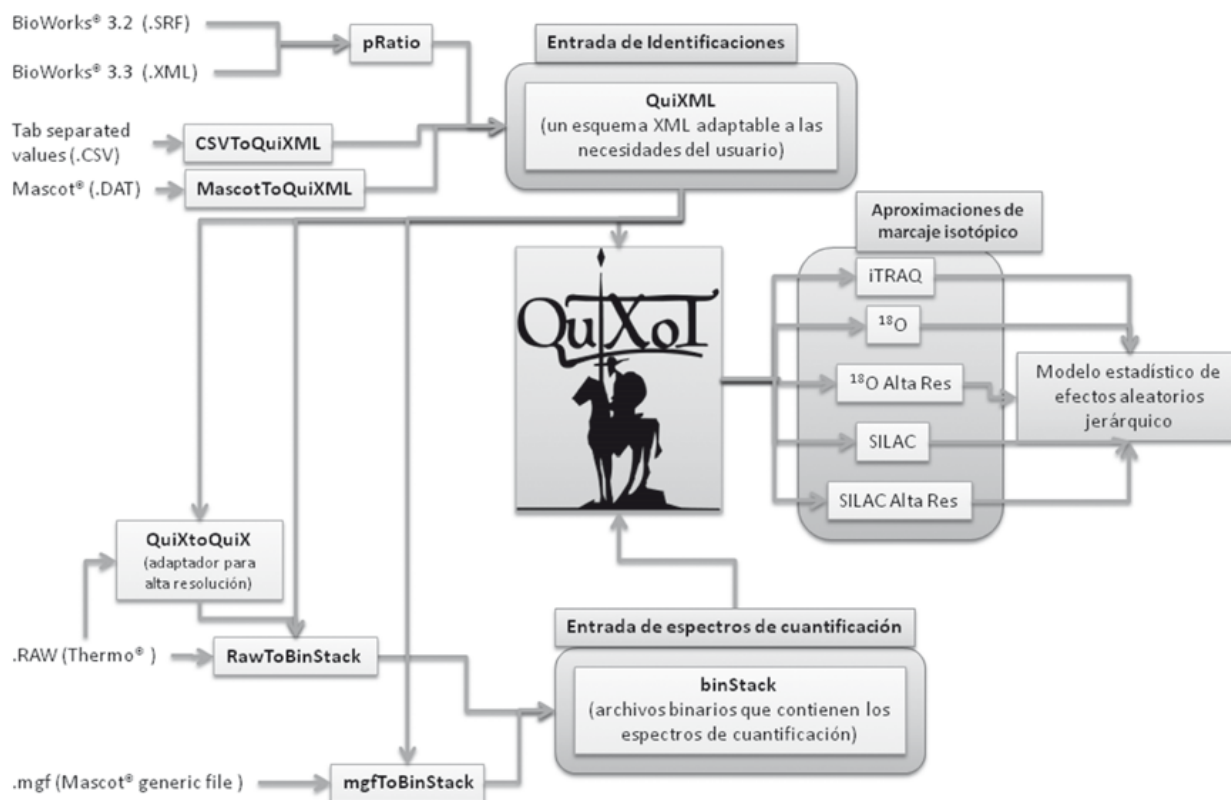
## Desarrollo de un Modelo Estadístico Universal para Proteómica Cuantitativa Mediante Marcaje Isotópico Estable

Marco Trevisan-Herraz<sup>1</sup>, Pedro Navarro<sup>1</sup>, Elena Bonzon-Kulichenko<sup>1</sup>, Pablo Martínez-Acedo<sup>1</sup>, Daniel Pérez-Hernández<sup>1</sup>, Estefanía Núñez<sup>1</sup>, María Luisa Hernández<sup>2</sup>, Enrique Calvo<sup>4</sup>, Montserrat Carrascal<sup>3</sup>, Marina Gay<sup>3</sup>, Inmaculada Jorge<sup>1</sup>, Dolores Gutiérrez<sup>2</sup>, Joaquín Abian<sup>3</sup>, Concha Gil<sup>2</sup>, Juan Miguel Redondo<sup>4</sup>, Jesús Vázquez<sup>1</sup>

<sup>1</sup>Laboratorio de Química de Proteínas y Proteómica, Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Madrid (CBMSO). <sup>2</sup>Departamento de Microbiología II, Facultad de Farmacia, Universidad Complutense de Madrid (UCM). <sup>3</sup>Laboratorio de Proteómica, Instituto de Investigaciones Biomédicas de Barcelona (IIBB-CSIC) (IDIBAPS). <sup>4</sup>Centro Nacional de Investigaciones Cardiovasculares, Madrid (CNIC)

Aunque en la actualidad existen diferentes métodos de marcaje isotópico para la realización de estudios de proteómica cuantitativa, los modelos estadísticos para el tratamiento de los datos obtenidos están todavía muy poco desarrollados. La mayoría de los modelos asumen distribuciones normales de los datos y la existencia de varianzas homogéneas, que no se han demostrado en la práctica. En el laboratorio de J. Vázquez hemos desarrollado recientemente un nuevo modelo estadístico jerárquico de efectos aleatorios para el análisis de los datos de expresión diferencial obtenidos por marcaje con <sup>18</sup>O/<sup>16</sup>O y espectrometría de masas de trampa iónica

lineal [1]. La validez de este modelo, que considera por separado las fuentes de error debidas a la manipulación de la muestra, la preparación de los péptidos y la cuantificación por el espectrómetro de masas, ha sido demostrada en experimentos masivos de hipótesis nula [1] y ha sido validada en varios modelos biológicos de diferente naturaleza [2]. En el presente trabajo hemos abordado un proyecto cooperativo a gran escala entre varios grupos de investigación con el objetivo de extrapolar dicho modelo estadístico, de forma general, a los métodos de marcaje isotópico más comunes (SILAC y iTRAQ), y a otros espectrómetros de masas.



**Figura 1.** Descripción del software que conforma la plataforma QuiXoT.

Como modelo biológico se ha empleado un único cultivo celular de *Saccharomyces cerevisiae*, control (muestra A), marcado con aminoácidos SILAC (A\*) o tratado con H<sub>2</sub>O<sub>2</sub> (B) (UCM), y se han realizado experimentos comparativos a gran escala A vs B y A vs A mediante <sup>18</sup>O/<sup>16</sup>O y iTRAQ, y A\* vs B y A\* vs A mediante SILAC. Las muestras se digirieron en el CBM utilizando un nuevo método de gel concentrante [2] y los péptidos resultantes se marcaron y separaron en 24 fracciones mediante la técnica Off-gel [2]. Las fracciones fueron analizadas usando un LTQ (CBM, IDIBAPS) en modo convencional o PQD, un LTQ-Orbitrap (CNIC) y un MALDI-TOF-TOF (UCM). Cada uno de los experimentos permitió la cuantificación de entre 3.000 y 5.000 péptidos. La eficiencia de marcaje por <sup>18</sup>O se determinó usando el método cinético [3]. Los resultados de los experimentos de hipótesis nula se utilizaron para el modelado de los pesos estadísticos y para el análisis de normalidad de las poblaciones a los diferentes niveles, según describimos previamente [1].

Nuestros resultados demuestran que el modelo estadístico jerárquico es válido, de forma universal, para todos los tipos de marcaje y todos los

espectrómetros de masas utilizados, obteniéndose distribuciones que superan los test de normalidad a nivel de medida, a nivel de péptido y a nivel de proteína. El modelo global ha sido implementado en la plataforma informática QuiXoT, en la que se han desarrollado interfaces para la entrada de datos obtenidos por los diferentes motores de búsqueda, equipos y métodos de marcaje (Figura 1).

El análisis de los datos arroja unas varianzas a nivel de péptido que son consistentemente semejantes para todos los métodos de marcaje post-digestión, mientras que en el caso de SILAC la varianza es prácticamente nula. Por otra parte, la varianza a nivel de medida en modo MS (SILAC, <sup>18</sup>O) es algo menor en el Orbitrap en relación con el LTQ, aumentando considerablemente en el modo PQD-MS/MS (iTRAQ) en el LTQ. Finalmente la varianza a nivel de proteína es muy baja y similar en todos los casos. Estos resultados demuestran la coherencia del modelo y establecen por primera vez un marco de referencia estadístico común para comparar las prestaciones relativas de los diferentes equipos y métodos de marcaje. Este modelo, por otra parte, devuelve información sobre la fiabilidad de los cambios de expresión obtenidos y ofrece un

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