

# 1. BIOINFORMÁTICA

Coordinadores: *Salvador Martínez-Bartolomé* y *Pedro Navarro*

## Mesa redonda sobre herramientas en bioinformática

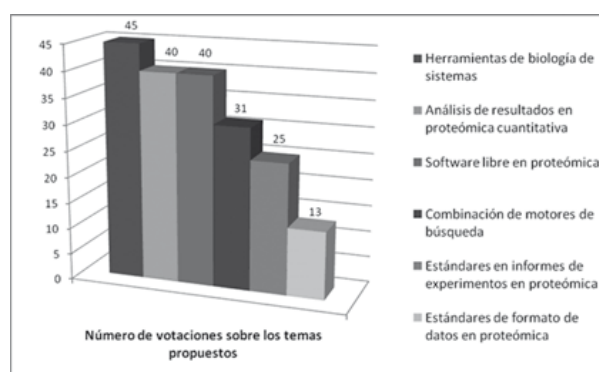
*Salvador Martínez-Bartolomé*<sup>1</sup>, *Pedro Navarro*<sup>2</sup>, *Alex Campos*<sup>3</sup>, *Marco Trevisan-Herraz*<sup>2</sup>, *Juan Antonio Vizcaíno*<sup>4</sup>, *Alberto Medina*<sup>1</sup>

<sup>1</sup>ProteoRed – Laboratorio de proteómica del Centro Nacional de Biotecnología – CSIC, Madrid. <sup>2</sup>Laboratorio de química de proteínas del Centro de Biología Molecular Severo Ochoa – CSIC, Madrid. <sup>3</sup>Plataforma de Proteómica, Parc Científic de Barcelona. <sup>4</sup>EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

El desarrollo y aparición de nuevas tecnologías aplicadas a la proteómica han dado lugar a un crecimiento espectacular de datos experimentales. La captura, el almacenamiento, el procesamiento y el análisis de estos datos es, en numerosas ocasiones, un cuello de botella que debemos solucionar para seguir avanzando en el campo.

Pese a que, paralelamente, han aumentado los proyectos de desarrollo de herramientas y servicios bioinformáticos aplicados a la proteómica, es muy difícil conocer todos los recursos disponibles. Muchos de estos proyectos son de software libre [<http://www.ms-utils.org/wiki/pmwiki.php/Main/Software-List>], y por tanto, disponibles de forma gratuita, la mayoría de las veces, a través de Internet. La aparición de la proteómica de segunda generación y los análisis a gran escala ha permitido los llamados estudios de biología de sistemas, cuyos análisis serían imposibles de hacer sin herramientas bioinformáticas adecuadas. Por otro lado, es destacable también el avance en los últimos años de nuevas técnicas de marcaje isotópico y label-free que permiten realizar estudios de proteómica cuantitativa. En dichos análisis, es necesario seguir metodologías estadísticas rigurosas para minimizar la detección de falsos cambios de expresión. Numerosos grupos de investigación se resisten a utilizar herramientas no comerciales, o desconocen la existencia de una alternativa. Otro problema existente relativo al manejo de los datos en proteómica es el intercambio de dichos datos entre diferentes plataformas, laboratorios y herramientas. Es por ello por lo que el desarrollo de estándares se hace crucial, y es necesario conocer las herramientas existentes que permiten la conversión entre los formatos propietarios y los estándares.

Todas estas cuestiones que surgen alrededor de la proteómica y la bioinformática nos han hecho pensar que un formato de mesa redonda podría ser ideal en unas jornadas que pretenden ser un espacio de intercambio de información entre los proteómicos y bioinformáticos españoles. En dicha mesa redonda pretendemos plantear los problemas que más nos preocupan a los proteómicos y que pueden ser resueltos, al menos en parte, por medio de recursos bioinformáticos. Por ello, se realizó una encuesta, cuyos resultados se muestran a continuación (Figura 1), para elegir de entre unos temas propuestos, los que resultan más interesantes para la comunidad científica proteómica. Una vez recopilados los prácticamente 200 votos que se recibieron, resultaron destacados tres temas principales: *herramientas de biología de sistemas*, *estadística en proteómica cuantitativa* y *software libre en proteómica*.



**Figura 1.** Número de votaciones obtenidas en la encuesta con la que se sondearon los temas más interesantes para la comunidad proteómica española. Los temas propuestos fueron: “herramientas de biología de sistemas”, “análisis de resultados en proteómica cuantitativa”, “software libre en proteómica”, “combinación de motores de búsqueda”, “estándares en informes de experimentos en proteómica” y “estándares de formato de datos en proteómica”.

Para la sesión hemos buscado personas con la suficiente experiencia y conocimientos como para ser las personas de referencia en el debate sobre dichos temas:

Alberto Medina (CNB-CSIC), pionero de la proteómica computacional en España, con dilatada experiencia en temas relacionados con la biología de sistemas como la búsqueda de anotaciones y los repositorios de información biológica y proteómica, así como en temas relacionados con estándares.

Alex Campos (PCB), experimentado usuario de herramientas de software libre aplicadas a la proteómica, así como de estándares y herramientas de conversión de formato de datos.

Marco Trevisan (CBMSO-CSIC), uno de los desarrolladores de un software de cuantificación de uso gratuito para instituciones públicas y conocedor de los problemas a resolver con la estadística en la cuantificación de proteínas.

Juan Antonio Vizcaíno (EMBL-EBI) integrante del proyecto PRIDE (*PRoteomics IDentifications*

*database*) [1], ha participado en estudios de minería de datos de espectrometría de masas y es participante en el desarrollo de estándares de la iniciativa de HUPO-PSI [2].

Cada uno de ellos realizará una muy pequeña introducción de los temas propuestos, y seguidamente se aprovechará la mayor parte del tiempo de la sesión en el intercambio de información entre los asistentes y nuestros “expertos”.

Creemos que entre los cuatro forman una mesa muy interesante y esperamos que podamos proporcionar una sesión fructífera para todos.

## Referencias

- [1] Vizcaino JA, Cote R, Reisinger F, Foster JM, Mueller M, Rameseder J, et al. A guide to the Proteomics Identifications Database proteomics data repository. *Proteomics* 2009;9:4276-83.
- [2] Orchard S, Hermjakob H y Apweiler R. The proteomics standards initiative. *Proteomics* 2003;3:1374-6.

## A new versatile file translator for Proteomics standards

*J. Alberto Medina-Aunon<sup>1,2</sup>, Salvador Martínez-Bartolomé<sup>1,2</sup>, J. Pablo Albar<sup>1,2</sup>*

<sup>1</sup>Proteomics Facility, National Center for Biotechnology (CNB-CSIC), <sup>2</sup>Spanish Proteomics Institute (ProteoRed).

Information derived by proteomics experiments such as sample preparation and separation, mass spectrometry, protein identification, etc..., is represented using different and sometimes incomprehensible file formats, resulting not very suitable for daily life. To deal with this problem, several initiatives have been developed: HUPO-PSI MIAPE and XML-based formats [1], XML-based repository: PRIDE [2], etc... All of them are contributing effectively within a Proteomics scope, but some lacks regarding the translation between some of the previous schemas are still confusing.

In order to contribute to this situation, we have developed a new Sun Java application accessible on

the Internet (<http://www.proteored.org/>). This tool has been designed using common graphics libraries and latest XML parsers offering a friendly and easy interface avoiding complex configuration dialogs.

Currently, the application is able to manage the following file formats -proteomics standards-: MIAPE Gel, MS and MSI documents (MS Excel and XML format), GelML, and PRIDE.

To illustrate this innovative work we realized two tests, both of them based on the translation between GelML and MIAPE Gel schemas. Input files were creating from 1) HUPO PSI Microsoft Excel template and 2) ProteoRed MIAPE web site ([www.proteored.org](http://www.proteored.org)). Both schemas are perfectly