

Desarrollo de herramientas bioinformáticas aplicadas a la identificación y cuantificación de péptidos en experimentos a gran escala

Navarro PJ¹, Jorge I¹, Martínez-Acedo P¹, Díaz M¹, Pérez D¹, Núñez E¹, Bonzón E¹, Serrano H^{1,2}, Vázquez J¹.

¹Centro de Biología Molecular Severo Ochoa, CSIC-UAM. ²Universidad de Puerto Rico en Arecibo, EE.UU.

Introducción

La identificación de péptidos a partir de espectros MS/MS y la cuantificación usando isótopos estables en experimentos de alto rendimiento requieren algoritmos de proceso de datos altamente automatizados. Nuestro laboratorio ha desarrollado herramientas para validar las identificaciones, pRatio (Martínez de Bartolomé et al, in review), y para calcular el ratio de expresión y la eficiencia de marcaje de los péptidos marcados mediante ¹⁸O, QuiXoT (Ramos et al., 2007). Recientemente hemos observado que el análisis de un experimento control a gran escala (hipótesis nula) mediante los modelos estadísticos convencionales da lugar a falsos cambios de expresión que resultan estadísticamente significativos. El objetivo de este trabajo es perfeccionar nuestras herramientas bioinformáticas actuales y la incorporación de un modelo estadístico adecuado.

Material y métodos

Los programas utilizados, pRatio y QuiXoT, han sido desarrollados íntegramente en nuestro laboratorio y han sido programado en C#.

Resultados

El experimento control es el análisis comparativo de un proteoma obtenido a partir de células endoteliales HUVEC contra sí mismo. Se identificaron más de 2700 péptidos, correspondientes a 1400 proteínas, de los que se consiguieron cuantificar más de 1500 péptidos correspondientes a más de 800 proteínas. Hemos introducido mejoras en pRatio que aumentan sus prestaciones, que superan las de otros métodos reconocidos publicados muy recientemente (Elias et al, 2007). También hemos desarrollado un modelo jerárquico lineal que ha permitido estudiar en detalle las diferentes

fuentes de error que influyen en los experimentos de cuantificación masiva mediante marcaje isotópico estable usando ¹⁸O. No se detectaron falsos cambios de expresión cuando el nuevo modelo se aplicó al análisis estadístico del experimento control.

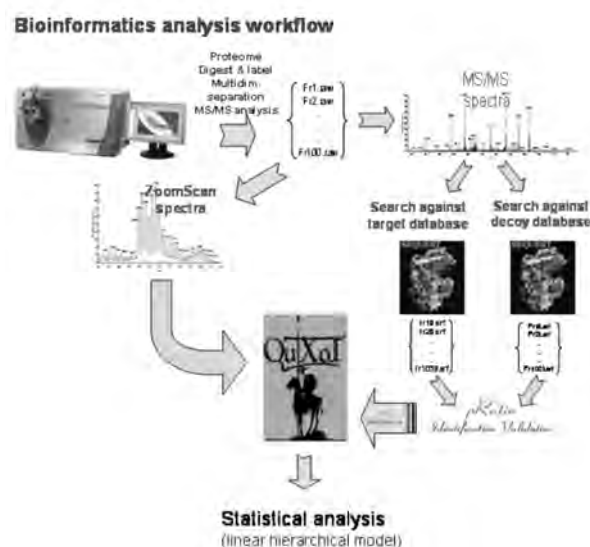


Figura 1: Se muestra el flujo del análisis de datos en un experimento de identificación y cuantificación de péptidos mediante marcaje isotópico estable de ¹⁸O. Los espectros MS/MS de los archivos correspondientes a las diferentes carreras del experimento fraccionado mediante intercambio catiónico (SCX) son buscados en una base de datos mediante un motor de búsqueda (SEQUEST). Los archivos de resultados de esta búsqueda son validados por el software pRatio, resultando en un único archivo de identificaciones válidas que es introducido en el programa de cuantificación (QuiXoT), el cual consulta los espectros ZoomScan asociados a cada identificación validada. Posteriormente se realiza el análisis estadístico que determina aquellos cambios de expresión que son estadísticamente significativos.

Conclusiones

Las herramientas bioinformáticas desarrolladas permiten un flujo de trabajo completo (Fi-

gura 1) y automatizado para la identificación y cuantificación masiva de péptidos mediante marcaje isotópico estable con ^{18}O , ofreciendo unas prestaciones superiores en cuanto a identificación (mayor número de identificados con la misma tasa de error) y en cuanto a robustez en la detección de cambios de expresión estadísticamente significativos.

La biología computacional de la proteómica post-genómica

Segura Ruiz V.

Unidad de Proteómica, Genómica y Bioinformática, CIMA

Introducción

El estudio de la expresión de genes mediante técnicas de alto rendimiento, como los microarrays de ADN, obligó a un cambio de paradigma en la interpretación de los resultados obtenidos con respecto a la experimentación clásica empleada en biología. La bioinformática tiene entre sus principales objetivos dar respuesta a estas nuevas necesidades desde múltiples disciplinas, como son la estadística y la informática. En proteómica las necesidades computacionales han surgido principalmente de los experimentos de identificación de proteínas mediante geles 2D y espectrometría de masas, es decir, han estado orientadas a la gestión de grandes cantidades de información, el procesado de imagen y de espectros y la búsqueda en bases de datos. Sin embargo, la aplicación de métodos ya empleados en genómica a estos datos, o la aparición de nuevas tecnologías como los microarrays de proteínas o el SELDI-TOF, abren nuevas expectativas en el área de la proteómica, tanto en el ámbito científico como en el ámbito clínico.

Material y métodos

El análisis de los datos obtenidos en un experimento de microarrays de ADN requiere una serie de procesos independientemente de la plataforma o del objetivo del estudio (Hoheisel, 2006): cálculo de los niveles de expresión utilizando el método de normalización adecuado, acondicionamiento de la matriz de expresión, y la aplicación del método de selección

Bibliografía

- Martinez-Bartolome, S., Martín-Maroto, F., Navarro, P.J., López-Ferrer, D., et al., *Mol. Cell. Proteomics* (in review).
- Ramos-Fernandez, A., Lopez-Ferrer, D. & Vazquez, J. *Mol Cell Proteomics*, 2007, 6, 1274-1286.
- Elias, J.E. & Gygi, S.P. *Nat Methods*, 2007, 4, 207-214.

(foldchange, estadística t, modelos lineales, etc), con el fin de identificar los genes de interés. A partir de esta lista de genes existe una gran diversidad de opciones para la interpretación de los resultados: análisis de los perfiles de expresión, análisis funcional, estudio de factores de transcripción, o la integración de información mediante redes de genes con programas como Ingenuity (www.ingenuity.com).

En el ámbito de la proteómica, con el empleo de nuevas tecnologías como los microarrays de anticuerpos (Madoz-Gúrpide et al, 2007) o el SELDI-TOF (Vorderwülbecke, 2005), el análisis precisa de métodos similares a los empleados en genómica, ya sea para el proceso de normalización, o para la selección, mediante métodos estadísticos y herramientas bioinformáticas apropiadas, de las proteínas alteradas o de la expresión diferencial de péptidos.

Conclusiones

La bioinformática permite, con datos provenientes de la experimentación en proteómica, el desarrollo de nuevas aplicaciones, como son la identificación de biomarcadores o de nuevas dianas terapéuticas para el desarrollo de medicamentos.

Bibliografía

- JD., Hoheisel. *Nature Reviews Genetics* 2006, 7, 3, 200-210.