

Métodos estadístico-computacionales para la caracterización de patrones de expresión de proteínas en 2D-DIGE.

Fernández, Elmer A. y Llera, Andrea Sabina y Balzarini, Mónica y Griotti, Romina Mónica y Ponzoni, Ignacio y Fresno, Cristóbal (2012) *Métodos estadístico-computacionales para la caracterización de patrones de expresión de proteínas en 2D-DIGE*. [Proyecto de Investigación]

El texto completo no está disponible en este repositorio. ([Solicitar una copia](#))

Resumen

El volumen de datos provenientes de experimentos basados en genómica y proteómica es grande y de estructura compleja. Solo a través de un análisis bioinformático/bioestadístico eficiente es posible identificar y caracterizar perfiles de expresión de genes y proteínas que se expresan en forma diferencial bajo distintas condiciones experimentales (CE). El objetivo principal es extender las capacidades computacionales y analíticas de los softwares disponibles de análisis de este tipo de datos, en especial para aquellos aplicables a datos de electroforesis bidimensional diferencial (2D-DIGE). En DIGE el método estadístico más usado es la prueba t de Student cuya aplicación presupone una única fuente de variación y el cumplimiento de ciertos supuestos distribucionales de los datos (como independencia y homogeneidad de varianzas), los cuales no siempre se cumplen en la práctica, pudiendo conllevar a errores en las estimaciones e inferencias de los efectos de interés. Los modelos Generalizados lineales mixtos (GLMM) permiten no solo incorporar los efectos que, se asume, afectan la variación de la respuesta sino que también modelan estructuras de covarianzas y de correlaciones más afines a las que se presentan en la realidad, liberando del supuesto de independencia y de normalidad. Estos modelos, más complejos en esencia, simplificarán el análisis debido a la modelización directa de los datos crudos sin la aplicación de transformaciones para lograr distribuciones más simétricas, produciendo también a una estimación estadísticamente más eficiente de los efectos presentes y por tanto a una detección más certera de los genes/proteínas involucrados en procesos biológicos de interés. La característica relevante de esta tecnología es que no se conoce a priori cuáles son las proteínas presentes. Estas son identificadas mediante otras técnicas más costosas una vez que se detectó un conjunto de manchas diferenciales sobre los geles 2DE. Por ende disminuir los falsos positivos es fundamental en la identificación de tales manchas ya que inducen a resultados erróneos y asociaciones biológica ficticias. Esto no solo se logrará mediante el desarrollo de técnicas de normalización que incorporen explícitamente las CE, sino también con el desarrollo de métodos que permitan salirse del supuesto de gaussianidad y evaluar otros supuestos distribucionales más adecuados para este tipo de datos. También, se desarrollarán técnicas de aprendizaje automática que mediante optimización de funciones de costo específicas nos permitan identificar el subconjunto de proteínas con mayor potencialidad diagnóstica. Este proyecto tiene un alto componente estadístico/bioinformática, pero creemos que es el campo de aplicación, es decir la genómica y la proteómica, los que más se beneficiarán con los resultados esperados. Para tal fin se utilizarán diversas bases de datos de distintos experimentos provistos por distintos centros de investigación nacionales e internacionales.

Tipología documental: Proyecto de Investigación

Palabras clave: Modelos lineales mixtos. Máquinas de soporte vectorial

Descriptor: [Q Ciencia > QA Matemáticas](#)

Unidad Académica: [Universidad Católica de Córdoba > Facultad de Ingeniería](#)