

(C-89)

**PLATAFORMA PARA EL APRENDIZAJE DE
ALGORITMOS COMPUTACIONALES Y SU
APLICACIÓN A LAS CIENCIAS DE LA VIDA.**

C. Cano

M. Cuadros

A. Labarga

A. Blanco



(C-89) PLATAFORMA PARA EL APRENDIZAJE DE ALGORITMOS COMPUTACIONALES Y SU APLICACIÓN A LAS CIENCIAS DE LA VIDA.

Autor/res/ras: *C. Cano, M. Cuadros, A. Labarga, A. Blanco*

Afiliación Institucional: Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada.

Indique uno o varios de los siete Temas de Interés Didáctico: (Poner x entre los [])

- Metodologías didácticas, elaboraciones de guías, planificaciones y materiales adaptados al EEES.
- Actividades para el desarrollo de trabajo en grupos, seguimiento del aprendizaje colaborativo y experiencias en tutorías.
- Desarrollo de contenidos multimedia, espacios virtuales de enseñanza- aprendizaje y redes sociales.
- Planificación e implantación de docencia en otros idiomas.
- Sistemas de coordinación y estrategias de enseñanza-aprendizaje.
- Desarrollo de las competencias profesionales mediante la experiencia en el aula y la investigación científica.
- Evaluación de competencias.

Resumen.

Con el avance de las nuevas tecnologías de alto rendimiento de procesamiento (high-throughput), la investigación en el campo de las Biociencias ha sufrido en las últimas décadas un cambio de rumbo por el enorme volumen de resultados biológicos obtenidos. Los resultados van desde el perfil de expresión génica, el análisis de polimorfismos, la creación de mapas metabólicos, la secuenciación de genomas completos, entre ellos el humano, hasta la integración e interpretación de dichos resultados para la cada vez más cercana “medicina personalizada”. Este tipo de investigaciones se realizan en centros de investigación especializados, hospitales, universidades y compañías farmacéuticas, y han producido un enorme aumento, en la última década, de la demanda de profesionales especializados en el análisis de la información biológica producida por estas tecnologías de alto rendimiento, en particular por la tecnología de Microarrays.

El EEES impone una nueva metodología en las enseñanzas universitarias en la que resulta fundamental incorporar las nuevas tecnologías e internet como vehículo de formación y experimentación. Este trabajo propone una herramienta web de apoyo a la docencia, tanto en el campo de las Biociencias (Biología, Bioquímica, Farmacia, Medicina, etc), como en el de las Tecnologías de la Información, que permita a los alumnos familiarizarse y experimentar con el análisis de datos biomédicos, en particular de Microarrays. Además, este proyecto pretende crear en los alumnos del campo de Tecnologías de la Información la motivación necesaria para aplicar sus conocimientos en el desarrollo de nuevas aplicaciones informáticas para el análisis e interpretación de este tipo de datos biológicos, contribuyendo a la especialización de estos alumnos hacia uno de los perfiles más demandado tanto en el mundo empresarial como en el de la investigación, el de bioinformático.

Keywords: Bioinformática, Biomedicina, Ciencias de la Computación, Microarrays

Abstract.

Scientific research in the Life Sciences domain has experienced a huge revolution with the advent of high-throughput technologies. These technologies are producing massive amounts of biomedical data, ranging from gene expression profiles to the sequencing on whole human genomes, enabling new important discoveries and allowing progress towards a personalized medicine. This data is produced and analyzed at research centres, hospitals, universities and pharmaceutical companies. Therefore, the need of highly-qualified professionals with a multidisciplinary background in Life Sciences and Computer Science, who are able to analyze these massive amounts of data and interpret the results from a biological perspective, has experienced an exponential raise in the last years. This paper discuss a novel perspective to incorporate Computer Science into Life Science degrees to allow this complementary formation.

The EEES imposes a new methodology in university degrees in which it is essential to incorporate the internet and the use of the new technologies as a vehicle for the formation and experimentation. This work presents a web-based tool that supports the teaching of expert algorithms and methods used to analyzed high-throughput biomedical data generated by Microarray Technologies. This web platform is to be used in both the Life Sciences domain (biology, biochemistry, Pharmacy, Medicine, etc.) and the Computer Science domain, and would allow the students to get used to the analysis of biomedical data from Microarrays. Furthermore, this project pretends to motivate the students from the Computer Science domain so they realize they can apply their knowledge in the development of new software for the analysis of biological data. We believe this contribution is essential for a bioinformatician, a professional profile which is increasingly demanded in the last year by the industry.

1. Introducción y objetivos

Tomando como punto de partida la importancia de la Bioinformática en el ámbito científico y partiendo del momento actual de convergencia hacia el Espacio Europeo de Educación Superior (EEES) [1], el trabajo que proponemos se enmarca y adecua a las nuevas técnicas docentes y planteamientos pedagógicos y metodológicos. Este trabajo presenta una metodología docente basada en la utilización de un portal web destinado al análisis de datos biomédicos, y en particular, de datos de expresión genética procedentes de microarrays. El objetivo del portal web es promover en los alumnos de Ingeniería Informática, Biología, Bioquímica, Farmacia y Medicina, entre otros, un aprendizaje activo de las nuevas tecnologías informáticas aplicadas a la biología molecular mediante el cual tomen conciencia de la importancia de la Bioinformática como herramienta fundamental dentro de su área de conocimiento. Para conseguir esta finalidad, el portal propicia la transversalidad entre las titulaciones de Informática, Biología, Bioquímica, Medicina y Farmacia.

1.1 Asignaturas de aplicación de la metodología propuesta

La metodología que se propone pretende propiciar la transversalidad entre las siguientes asignaturas impartidas por distintos departamentos de la Universidad de Granada:

- Estructura de Datos, Teoría de Algoritmos y Neuro-computación, impartidas por el Dpto. Ciencias de la Computación en la titulación de Ingeniería Informática.
- Bioinformática e Informática aplicada a la Bioquímica, impartidas por el Dpto. Ciencias de la Computación en las titulaciones de Biología y Bioquímica, respectivamente.

- Bioquímica y Biología Molecular, impartida por el Dpto. Bioquímica y Biología Molecular en la titulación de Farmacia.

1.2 Objetivos

Los objetivos específicos de la metodología que proponemos son los siguientes:

- Fomentar un modelo de enseñanza basado en la transversalidad de los contenidos y el aprendizaje autónomo y en equipo por parte de los estudiantes.
- Permitir a los estudiantes el acceso a un servidor web para la ejecución en remoto de software de análisis de datos de expresión genética que les permita realizar prácticas con datos reales e iniciarse en algunas de las tareas típicas de la bioinformática.

Los beneficios potenciales de estas actividades se enumeran a continuación:

- Mostrar a los alumnos de enseñanzas técnicas las posibilidades que ofrece la aplicación de los conocimientos teóricos adquiridos sobre problemas reales de actualidad, y en particular, sobre la extracción de información a partir de datos biomédicos. De este modo, se les muestra la importancia y las posibilidades de la interdisciplinariedad en estos campos, y se les inicia en un campo falto de profesionales a nivel nacional.
- Ampliar el campo de salidas profesionales, proporcionándoles una formación adecuada en el uso de técnicas actuales aplicadas al campo de las Biociencias, mejorando la adquisición de competencias profesionales en el mundo laboral y posibilitando la relación teoría-práctica.
- Permitir que estudiantes del campo de las Biociencias (Biología, Química, Genética, Bioquímica, Farmacia, Medicina, etc) puedan realizar prácticas on-line sobre problemas reales, utilizando técnicas de última generación y potentes herramientas bioinformáticas, que les permita adquirir destrezas necesarias para el desarrollo de parte de las actividades de su futuro profesional.

Cumplir con estos objetivos permitirá centrar al alumno como un elemento activo del proceso de enseñanza y aprendizaje, aplicando los contenidos teóricos en asignaturas involucradas para la resolución de problemas reales.

2. Descripción de la plataforma web

El proyecto se basa en una herramienta web que permite la ejecución remota de tecnologías Bioinformáticas. El proyecto persigue ofrecer al estudiante una formación actualizada incorporando herramientas Bioinformáticas para el análisis de microarrays. Los alumnos utilizarían el portal web para analizar datos de expresión genética, siguiendo simulaciones y guiones desarrollados por el equipo docente. La implantación de este proyecto está programada para el curso académico 2010-2011.

De una manera esquemática, el portal se puede dividir en dos módulos generales:

- 1.- Gestión de usuarios y de contenidos de cada usuario.
- 2.- Plataforma software para análisis de microarrays.

El primer módulo general permite a los usuarios registrarse y mantener una cuenta con sus propios ficheros de datos. Además, se pone a disposición del usuario el acceso a microarrays de ejemplo procedentes de bases de datos públicas. El usuario puede monitorizar la lista de tareas en ejecución y crear/abortar nuevas tareas.

El segundo módulo general pone a disposición de los alumnos distintos programas para realizar un análisis completo de microarrays de expresión de distintas plataformas comerciales, desde el preprocesamiento y normalización de los datos, hasta la interpretación biológica de los resultados. Este módulo se compone, a su vez, de los siguientes elementos:

2.1. Análisis de la calidad de los arrays. Para la evaluación de la calidad de los microarrays se ponen a disposición de los alumnos distintas herramientas de fuerte base estadística, basadas en publicaciones de reciente aparición en la literatura especializada, como el paquete de BioConductor `arrayQualityMetrics` [2]. Estas herramientas han sido incorporadas a programas implementados en lenguaje R con los que se generan informes con diversos tipos de gráficas que permiten evaluar la calidad individual de cada array, la reproducibilidad de los experimentos, la homogeneidad entre experimentos y la evaluación del ratio señal/ruido.

2.2. Pre-procesamiento. Este paso consiste fundamentalmente en estimar y eliminar los efectos que son debidos a las variaciones en la tecnología y no a la biología. La plataforma pone a disposición de los alumnos los métodos de normalización por mediana, RMA, GCRMA, MAS, PLIER [3].

2.3. Análisis estadístico de datos de expresión genética. Este módulo dispone de distintos componentes para el análisis de datos de expresión genética con distintos métodos estadísticos (Significance Analysis of Microarrays, SAM [4]), para la determinación de genes con perfiles de expresión significativamente diferentes entre subgrupos de muestras proporcionados por el usuario.

2.4. Algoritmos de Clustering. Este módulo dispone de distintos componentes para el análisis no supervisado de los datos de expresión para la identificación de clusters y biclusters que relacionen genes co-expresados para distintas muestras en estudio [5].

2.5. Enriquecimiento semántico de los resultados. Este módulo dispone de distintos componentes de ayuda a la interpretación biológica de los resultados obtenidos de la ejecución del software en 2.3 y 2.4 mediante el estudio del enriquecimiento estadístico de los grupos de genes obtenidos en categorías de bio-ontologías (Gene Ontology [6], KEGG [7]).

2.6. Consulta en otras aplicaciones bioinformáticas. Este módulo ofrece componentes de consulta en otras suites de software para análisis de datos de expresión como FATIGO [8] o STRING [9]. Del mismo modo, se dispone de componentes de visualización y navegación de los resultados obtenidos, incluyendo gráficas de expresión para los genes de los agrupamientos obtenidos y redes de regulación que muestren las conexiones entre grupos de genes de interés.

3. Conclusiones

En este trabajo se han conseguido dos objetivos. Por una parte se ha puesto en marcha una plataforma web que integra distintos algoritmos y herramientas bioinformáticas para el análisis de datos de microarrays. Por otra parte, esta nueva plataforma está permitiendo crear material docente de prácticas para alumnos de un amplio abanico de titulaciones de los campos de las Biociencias y las Tecnologías de la Información que les permite afrontar problemas reales de análisis de datos biomédicos desde una perspectiva técnica y computacional. Esta plataforma y material docente ha comenzado a utilizarse en distintas asignaturas en el curso académico 2010-2011.

Agradecimientos

El trabajo de MC, AB y CC está financiado por los proyectos P08-TIC-4299 de la Junta de Andalucía y TIN 2009-13489 del Ministerio de Ciencia.

Bibliografía y Referencias.

- [1] Ministerio de Ciencia e Innovación, Espacio de Educación Superior Europeo, <http://web.micinn.es/>.
- [2] Kauffmann A, et al. (2008) A bioconductor package for quality assessment of microarray data. *Bioinformatics* 25(3):415-416.
- [3] Seo J, Hoffman EP (2006) Probe set algorithms: is there a rational best bet? *BMC Bioinformatics* 7:395
- [4] Tusher S, et al. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *PNAS* 98: 5116-5121.
- [5] Cano C, et al. (2009) Intelligent System for the Analysis of Microarray Data Using Principal Components and EDAs. *Expert Systems with Applications* 36(3): 4654-4663.
- [6] The gene ontology. <http://www.geneontology.org/>
- [7] KEGG: Kyoto Encyclopedia of Genes and Genomes. <http://www.genome.jp/kegg/>
- [8] Al-Shahrour F, et al. (2004) FatiGO: a web tool for finding significant associations of GO terms with groups of genes. *Bioinformatics* 20,4.
- [9] Jensen LJ, et al. (2009) STRING 8. *Nucleic Acids Res.* 37:D412-6.

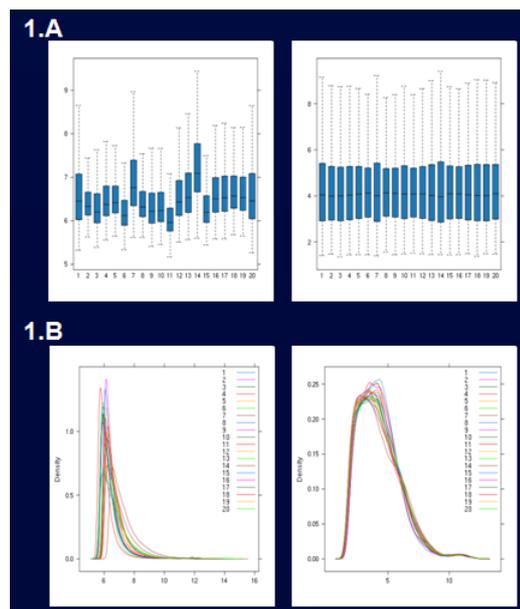


Figura 1. 1.A). Diagrama de cajas del log2 (Intensidades). Cada caja se corresponde con un array. Podemos comprobar que tras la normalización los microarrays son homogéneos ya que las cajas son de amplitud similar y están centradas en posiciones cercanas de el eje Y; Figura 1.B). Representación gráfica de la calidad del material biológico, en este caso ARN, empleado. Cada línea de corresponde con una muestra hibridada en un microarray.

Figura 2. Esquema general del proceso de obtención y análisis de los datos de microarrays

