Actas de las I Jornadas sobre Innovación Docente y Adaptación al EEES
en las Titulaciones Técnicas, Granada 2010, pp. 53-56
ISBN 978-84-92757-64-0 DL GR 3336-2010
Ed. Godel Impresores Digitales S.L.

RECURSO WEB PARA EL APRENDIZAJE DE APLICACIONES BIOINFORMÁTICAS: ANÁLISIS DE MICROARRAYS.

Cuadros Marta [(1)], Cano Carlos [(1)], Armando Blanco [(1)]

[(1)] *Departamento de Ciencias de la Computación e Inteligencia Artificial. Escuela Técnica Superior de Ingeniería Informática y de Telecomunicación. Universidad de Granada.*

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".
Gran parte de las investigaciones que se realizan a nivel de las Biociencias en Centros de Investigación, Hospitales, Universidades, Compañías Farmacéuticas etc., incluyen entre sus protocolos la producción y análisis de *Microarrays*, que en la mayoría de los casos son solicitados a empresas especializadas. Sin embargo el análisis que ofrecen dichas empresas es demasiado general, no atendiendo a las peculiaridades de los mismos. Ésta situación ha motivado un aumento drástico a nivel mundial de las ofertas de trabajo para profesionales formados en el análisis de datos biomédicos.
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.

Palabras clave

Microarrays, Bioinformática

1. INTRODUCCIÓN Y OBJETIVOS

Partiendo del momento actual de convergencia hacia el Espacio Europeo de Educación Superior (EEES) [1] y tomando como punto de partida la importancia de la Bioinformática en el ámbito científico, 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 donde se aplicará la metodología

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 Depto. Ciencias de la Computación en la titulación de Ingeniería Informática.
- Bioinformática e Informática aplicada a la Bioquímica, impartidas por el Depto Ciencias de la Computación en las titulaciones de Biología y Bioquímica, respectivamente.
- Bioquímica y Biología Molecular, impartida por el Depto. 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 a la práctica de las asignaturas involucradas ya que usarán el conocimiento teórico para la realización de una práctica real. El alumno adquirirá la capacidad de resolver problemas reales, que cada vez serán más cotidianos en cualquier trabajo relacionado con las diversas áreas de la Biología, Bioquímica, Farmacia, Medicina, etc., aumentando el interés del alumno por los contenidos de las asignaturas afectadas.

Con la metodología y los recursos didácticos puestos a disposición del alumnado a través de esta aplicación web, no pretendemos cubrir por sí mismo todo el proceso de formación sino complementar los estudios teóricos (en el caso de alumnos del campo de las Biociencias) con prácticas reales que acerquen a los alumnos a nuevas habilidades técnicas que serán de gran utilidad en su desarrollo profesional, y en el caso de los

alumnos de áreas técnicas, motivarlos para que apliquen y desarrrollen sus conocimientos en el área de la bioinformática.

2. DESCRIPCIÓN DEL PORTAL 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 (Affymetrix, Agilent, Codelink), 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:

- **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 basadas en publicaciones de reciente aparición en la literatura especializada, como el paquete de BioConductor arrayQualityMetrics [2]. Utilizando estas herramientas se han implementado programas en lenguaje R que generan informes con diversos tipos de gráficas que permiten evaluar la calidad individual de cada array, la existencia de efectos espaciales de error en los arrays, la reproducibilidad de los experimentos, la homogeneidad entre experimentos y la evaluación del ratio señal/ruido.
- **Pre-procesamiento.** El pre-procesamiento de los datos de expresión incluye por lo general diversas tareas, tales como la normalización, relativización de los valores de expresión, tratamiento de datos perdidos y eliminación de datos planos. La normalización es el más importante de estos pasos, y consiste en 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, background + loess [3].
- **Análisis estadístico de datos de expresión genética.** Este módulo dispone de distintos componentes para el análisis estadístico 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.
- **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. Se han implementado distintos algoritmos de clustering ampliamente utilizados para el análisis de microarrays (K-medias, Gene-Shaving) [5], así como de biclustering (Gene-&-Sample Shaving, EDA-Biclustering) [6].

- **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 [7], KEGG [8]).
- **Consulta en otras aplicaciones bioinformáticas.** Este módulo ofrece componentes de consulta en otras suites de software para análisis de datos de expresión como FATIGO [9], DAVID [10] o STRING [11]. 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. Este módulo también incluye software de extracción de reglas de asociación, que permite identificar patrones frecuentes y significativos que relacionen valores de expresión y otras variables de interés (propiedades de las muestras, tiempo de supervivencia, valores de variables de diagnóstico, etc.).

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 comenzará 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

[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] Hastie T, *et al.* (2000) 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. Genome Biology 2000, 1:3

[6] 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.

[7] The gene ontology. http://www.geneontology.org/

[8] KEGG: Kyoto Encyclopedia of Genes and Genomes. http://www.genome.jp/kegg/

[9] Al-Shahrour F, *et al.* (2004) FatiGO: a web tool for finding significant associations of GO terms with groups of genes. Bioinformatics 20,4.

[10] Huang DW, *et al*. (2009) Systematic and integrative analysis of large gene lists using DAVID. Nature Protoc. 4(1):44 -57.

[11] Jensen LJ, *et al*. (2009) STRING 8. Nucleic Acids Res. 37:D412-6.