

Plataforma Integral de Análisis de Microarreglos de Affymetrix: Aplicación al Estudio de la Farmacología del Parkinson

González G¹, Larramendy C^{2,3}, Gershanik O^{2,3}, Fernández E^{1,2}

**¹Grupo de Minería de Datos en Biociencias, Universidad Católica de Córdoba,
Argentina; ²Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina;**

³Laboratorio de Parkinsonismo Experimental, UBA-Conicet, Argentina.

Resumen

Las tecnologías de alto rendimiento empleadas actualmente en los laboratorios de biología molecular generan una gran cantidad de datos, que hacen necesaria la utilización de herramientas informáticas para su análisis. Estas herramientas requieren de diversas habilidades e infraestructura para su utilización, no siempre disponibles en los equipos de investigación de Argentina. Por ello, en este trabajo proponemos una plataforma de almacenamiento de experimentos de microarreglos de ADN y una rutina de procesamiento que automatiza los procesos de evaluación de la calidad y análisis de experimentos. La misma fue aplicada sobre un experimento de expresión génica en Parkinson.

Palabras claves: Minería de datos, microarreglos de ADN, pipeline.

Introducción

A pesar de que han pasado más de 10 años de la secuenciación del genoma humano, solo se conoce la función de una parte muy pequeña de los genes que lo componen. Allí es donde entra en juego la tecnología de microarreglos que permite monitorear simultáneamente la expresión de miles de genes. Conociendo la diferencia en el nivel de expresión bajo diferentes condiciones es posible hipotetizar acerca de la función de un gen dado o de su participación en un determinado proceso biológico.

Debido a la gran cantidad de datos que se generan en un experimento de este tipo, la bioinformática se convierte en una herramienta crucial para su manipulación y análisis. Por ello se han desarrollado numerosas aplicaciones, que intentan brindar soluciones apropiadas. El problema radica en decidir cuál de estas herramientas es la indicada para el grupo de trabajo que la va a utilizar. La situación de un laboratorio de biología molecular en Argentina presenta algunas particularidades que deben ser tenidas en cuenta a la hora de decidir esto. Por un lado gran parte de los programas disponibles actualmente para realizar este tipo de análisis tienen una curva de aprendizaje alta por lo que su utilización se torna dificultosa por parte de los usuarios que no tienen conocimientos avanzados de informática. Otro factor importante a tener en cuenta es que lado la mayoría de estos laboratorios no cuentan ni con un bioinformático ni con una plataforma informática adecuada que le permita administrar, almacenar y analizar experimentos de microarreglos. Por esta razón, se ven obligados a recurrir a grupos externos que realicen esta labor o a utilizar herramientas disponibles en Internet que no siempre satisfacen las necesidades específicas de estos grupos de investigación.

En este trabajo se propone abordar esta carencia mediante la implementación de una plataforma bioinformática flexible y adaptativa para el almacenamiento y análisis de experimentos de microarreglos de ADN. Dicha plataforma se basa en la utilización de las herramientas que brinda el proyecto caBIG, para la administración y almacenamiento de los datos producidos por esta tecnología. Estas herramientas permiten automatizar procesos

analíticos mediante el desarrollo de rutinas o *pipelines* de procesamiento que guardan las funciones y los parámetros elegidos por el usuario. También es posible extender sus funcionalidades mediante la creación de módulos que realizan tareas específicas. En este trabajo, haciendo uso de esta característica, se desarrolló un módulo de software para el análisis de experimentos tipo caso-control en microarreglos de ADN de la empresa Affymetrix. Este módulo permite realizar una tarea que no era realizada por ningún otro módulo dentro del sistema y que es importante a la hora de analizar microarreglos. La confluencia de dicho módulo con las características de caBIG permite, además de realizar análisis estadísticos y de clasificación, integrar estos datos experimentales con aquellos provenientes de estudios clínicos.

Estas herramientas serán aplicadas sobre datos provenientes de un experimento de expresión génica realizado en el Laboratorio de Parkinsonismo Experimental (UBA-Conicet). El objetivo de este experimento es conocer el perfil de expresión inducido por dos fármacos muy utilizados en el tratamiento del Parkinson, levodopa y pramipexole.

Elementos del trabajo y metodología

La plataforma propuesta en este trabajo se basa en las herramientas provistas por el proyecto caBIG (Cancer Biomedical Informatics Grid). Estas herramientas desarrolladas y mantenidas por el National Cancer Institute (NCI) de los Estados Unidos tienen como objetivo facilitar la administración de proyectos de salud, compartiendo información entre los diversos integrantes de la comunidad médica y científica [<https://cabig.nci.nih.gov/overview/>]. En el área molecular, que es en la que éste trabajo hará énfasis, lo que se busca es simplificar el almacenamiento de los datos de microarreglos; permitir la integración, visualización y análisis; y facilitar la gestión de los resultados.

Para poder cumplir con estos objetivos, caBIG cuenta con diversas aplicaciones especializadas. Este trabajo se centra en instalar, adaptar y configurar tres de ellas: caArray, caIntegrator2 y GenePattern. En la figura 1 se puede ver como se relacionan estas tres aplicaciones dentro del esquema propuesto, siendo caArray el “almacén” de datos, GenePattern la herramienta de análisis (pudiendo usar funciones de Bioconductor) y caIntegrator la plataforma que integra todos estos datos entre sí y con los datos clínicos.

Mediante estas aplicaciones se buscar analizar datos provenientes de análisis de microarreglos de la empresa Affymetrix, los cuales consisten básicamente en las siguientes etapas:

- Lectura de datos
- Control de calidad
- Filtrado
- Corrección del ruido de fondo
- Normalización
- Estimación del valor de expresión
- Análisis de expresión diferencial

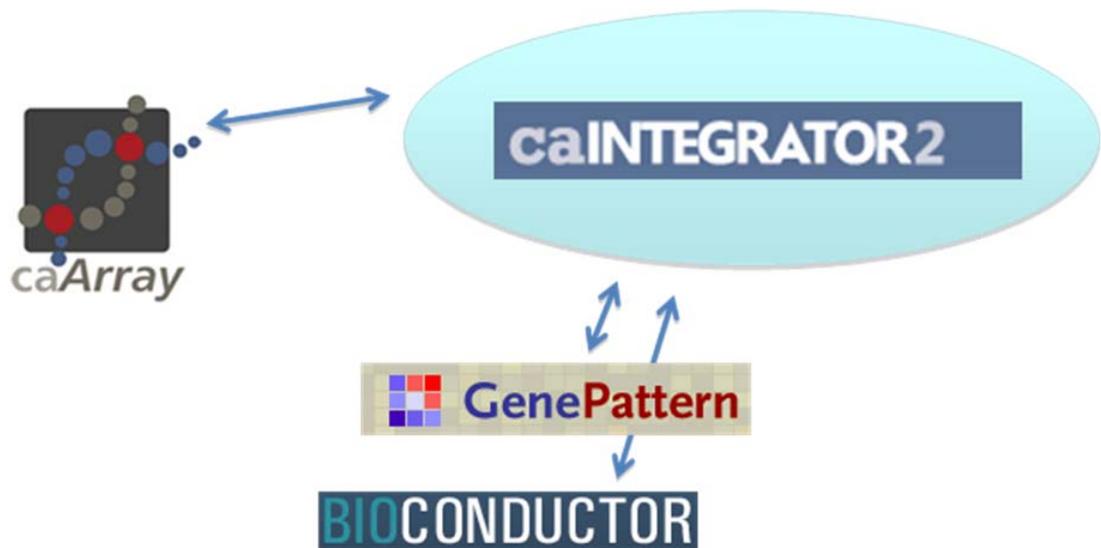


Figura 1. Integración de las herramientas caArray, GenePattern y caIntegrator2.
Adaptado de [1].

Existe una gran cantidad de herramientas que permiten realizar este tipo de análisis, siendo R una de las más utilizadas. Si bien R cuenta con una gran cantidad de funciones para el análisis de datos genómicos, su curva de aprendizaje es muy alta. Para utilizarlo el usuario debe aprender un lenguaje de programación con todas las dificultades que eso trae aparejado, especialmente para usuarios que no vienen del ámbito informático. Mientras que la plataforma propuesta en este trabajo permite realizar estas tareas mediante una sencilla interfaz web desde donde se pueden realizar gran parte de las tareas de análisis sin necesidad de tener conocimientos de programación. Además, es posible crear rutinas que automatizan gran parte del proceso de análisis reduciendo en gran medida el tiempo necesario para realizar estas tareas.

Flujo de trabajo en la plataforma

La primera etapa de este análisis consiste en cargar los datos en la plataforma caArray para que luego puedan ser analizados utilizando las herramientas provistas por GenePattern. caArray cuenta con una intuitiva interfaz de usuario donde se cargan los archivos .CEL y se agrega toda la información relevante que está relacionada con el experimento (figura 2). Esto es importante ya que permite centralizar los datos referentes a cuáles son los experimentos que realiza un laboratorio, cuáles fueron las condiciones bajo análisis y qué tipo de muestras se utilizaron. Teniendo esta información debidamente registrada se evita el desperdicio de insumos debido a experimentos duplicados.

Experiment: Estudio farmacogenómico de Parkinson

The screenshot shows the GenePattern interface for an experiment titled "Estudio farmacogenómico de Parkinson". At the top, there are tabs for "Overview", "Contacts", "Annotations", "Data", and "Publications". Below these, there are sub-tabs for "Manage Data", "Imported Data", "Supplemental Files", and "Download Data". The "Imported Data" sub-tab is active, showing a table of imported files. The table has a filter dropdown set to "(All)". The table columns are: File Name, File Type, Status, Compressed Size, and Uncompressed Size. There are 12 rows of data, each representing a different CEL file imported from Affymetrix.

<input type="checkbox"/>	File Name	File Type	Status	Compressed Size	Uncompressed Size
<input type="checkbox"/>	A (NA1) 1_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	B (LA1) 2_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	C (LL1) 3_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	D (LP1) 4_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	E (NA2) 5_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	F (LA2) 6_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	G (LL2) 7_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.1 MB	10.6 MB
<input type="checkbox"/>	H (LP2) 8_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	I (NA3) 9_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	J (LA3) 10_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	K (LL3) 11_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.2 MB	10.6 MB
<input type="checkbox"/>	L (LP3) 12_(RaGene-1_0-st-v1).CEL	Affymetrix CEL	✓ Imported	4.3 MB	10.6 MB

Figura 2. Tabla dentro de caArray que muestra los archivos de expresión (.CEL) importados.

Para poder realizar los análisis posteriores es necesario importar estos datos dentro de GenePattern mediante el módulo CaArray2.3.0Importer. Allí se debe especificar la URL en la que funciona caArray, el nombre del experimento que se quiere importar y los datos necesarios para iniciar sesión en caArray (ver figura 3).

La etapa de control de calidad es llevada a cabo por el módulo MadMax Array Quality Control creado por NuGO (*Nutrigenomics Organisation* [2]). Este módulo brinda una gran cantidad de datos útiles a la hora de evaluar la calidad de los chips que se están usando, como pueden ser gráficos de caja, histogramas, visualización de los datos crudos, gráficos de correlación y gráficos M vs. A. La mayor ventaja con respecto a otros sistemas es que permite realizar todas estas tareas de una forma simple y unificada. El usuario sólo debe cargar los datos que quiere analizar y el módulo devuelve como resultado un archivo comprimido que contiene todos los gráficos de calidad mencionados anteriormente.

Además de contar con el CDF proporcionado por Affymetrix, GenePattern cuenta con la opción de utilizar CDFs personalizados como los que desarrolla el Molecular and Behavioral Neuroscience Institute (MBNI). La diferencia radica en que las sondas son reorganizadas de acuerdo a la última información del genoma y del transcriptoma [http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic_curated_CDF.asp]. Algunos trabajos científicos han demostrado que esta actualización de la información aumenta tanto la precisión como la exactitud de los análisis [3] [4]. Se han encontrado diferencias de entre un 30 y un 50% en la lista de genes diferenciales obtenida con los CDFs originales y con los personalizados [5].

CaArray2.3.0Importer version 3

* required field Run Reset properties | export | help

url*
URL for caArray server.

experiment*
Title or Identifier of the caArray experiment to import data from.

type* ▾
Type of bioassay data to be retrieved.

extension
Data file extension to be retrieved.

zipFileName*
Name of the zip file to be created.

username
Username for caArray server

password
Password for caArray server

Run Reset properties | export | help

Figura 3. Importación de datos dentro de GenePattern utilizando caArrayImporter.

En el caso de que se utilice el CDF de MNBI no es necesario realizar un filtrado posterior de los genes de control. Si en cambio se usa el CDF de Affymetrix, se torna inevitable filtrar los genes de control. El problema radica en que a pesar de que se han desarrollado más de 100 herramientas para GenePattern no existe un módulo en que permita filtrar las sondas para seleccionar solo aquellas propias del experimento. Esto es una muestra de la gran variedad de problemas que pueden surgir durante un análisis de microarreglos y de la necesidad de utilizar herramientas flexibles que permitan agregar nuevas funcionalidades en caso de ser necesario. Para poder resolver este problema se desarrolló un módulo propio que realizado el filtrado de los datos. Debido a la versatilidad de GenePattern es posible programar funciones en cualquier lenguaje y luego importarlas dentro de la plataforma, lo que permitió escribir una función en R para realizar dicha tarea. Una vez que la función es importada, la aplicación permite crear una interfaz web de una manera simple e intuitiva. El módulo es muy sencillo de utilizar, como se puede ver en la figura 4. Solo se deben cargar el archivo con los datos preprocesados (que en esa aplicación tienen la extensión GCT) y un archivo de anotación con extensión TAB que es provisto por Affymetrix. Con estos datos la función filtra el archivo y lo devuelve en otro con la misma extensión.

Para la etapa de preprocesamiento, que incluye la corrección de ruido de fondo, la normalización y la estimación del valor de expresión se utilizó el algoritmo RMA (Robust Media Average [6]). Este algoritmo se encuentra implementado dentro del módulo *NuGO Expression File Creator*. Un parámetro importante que se debe definir antes de ejecutar esta función es cuál es el CDF que se utilizará, si el de Affymetrix o el del MNBI.

Por último, el análisis de expresión diferencial se realiza utilizando el módulo Limma Analysis [7]. Como resultado el módulo devuelve un archivo por cada comparación que incluyen la anotación de cada gen.

Figura 4. Interfaz del módulo PreprocessAffy, creado en este trabajo.

Los datos genómicos generados pueden ser relacionados con la información del mismo paciente que se encuentra esparcida por diferentes sistemas informáticos utilizando el software caIntegrator2. Esto la convierte en una herramienta importante para administrar estudios en humanos que involucren experimentos genómicos, estudios clínicos e imágenes biomédicas. Al igual que GenePattern, caIntegrator2 puede comunicarse con caArray para administrar los datos de microarreglos. Esto se logra mediante un módulo muy similar al que se utiliza en GenePattern.

Rutina automatizada

Los pasos que se han mencionado en la sección anterior pueden ser automatizados de forma tal que en futuros análisis que utilicen la misma tecnología solo sea necesario cargar los archivos y se obvian las tareas de elegir las herramientas y los parámetros. Esto es lo que se conoce como rutinas o *pipelines* en GenePattern.

La rutina creada consiste en tres etapas: en primer lugar el control de calidad provisto por MADMAX Array Quality Control, luego el preprocesamiento de los datos utilizando NuGO Expression File Creator y por último el análisis estadístico mediante el módulo Limma Analysis.

Para utilizar la rutina sólo es necesario cargar el contenedor ZIP que almacena los archivos CEL del experimento, el archivo CLS que detalla las condiciones hibridadas en cada chip e introducir una dirección de correo electrónico en caso de que quiera ser notificado cuando los resultados estén listos (figura 5).

Figura 5. Interfaz de la rutina creada

Aplicación sobre un estudio de la farmacología del Parkinson

En el experimento analizado se busca comparar el perfil de expresión de transcritos inducido por tratamiento con dos fármacos muy utilizados levodopa y pramipexole.

El modelo animal utilizado consiste en ratas Wistar macho a las cuales se les realiza una lesión en el cerebro. Luego se seleccionan aquellas ratas que muestran deficiencias de comportamiento claramente observables y se dividen en tres grupos para estudiar el efecto de los fármacos, uno recibe levodopa, otro pramipexole y el último solamente agua. Luego de tres semanas de tratamiento se extrae el cerebro de la rata y se toma el ARN total para llevar a cabo un análisis de los perfiles de transcripción utilizando microarreglos RaGene 1.0 ST de Affymetrix.

El experimento consiste en 12 microarreglos que contienen un pool de cuatro ratas cada uno. Los chips se encuentran divididos en cuatro condiciones experimentales, de las cuales se han realizado tres réplicas (ver Tabla 1). La condición NA ("normal agua") se corresponde con un grupo de control que incluye a ratas que solo tomaron agua durante el tratamiento. El grupo LA ("lesionado agua") corresponde a ratas lesionadas que tomaron agua, es decir que, se trata de otro grupo de control. El grupo LL ("lesionado levodopa") está formado por ratas que tomaron el fármaco levodopa. Y por último se encuentra el grupo LP ("lesionado pramipexole"), en el cual se encuentran las ratas lesionadas que han recibido dicha droga.

Condiciones experimentales	Réplicas
NA	A (NA1)
	E (NA2)
	I (NA3)
LA	B (LA1)
	F (LA2)
	J (LA3)
LL	C (LL1)
	G (LL2)
	K (LL3)
LP	D (LP1)
	H (LP2)
	L (LP3)

Tabla 1. Diseño del experimento

Resultados

Los archivos correspondientes a los 12 chips fueron cargados en caArray donde también se incluyó la información acerca del diseño del experimento y el origen de las muestras. Luego estos archivos fueron importados dentro de GenePattern y cargados en la rutina desarrollada en este trabajo. Los resultados de cada etapa son mostrados en una página web (figura 6) desde donde pueden ser descargados a su computadora local.

En el último paso se obtienen las seis tablas creadas por Limma con las diferentes comparaciones. Las listas son filtradas por valor p y por fold change para conservar únicamente aquellos transcritos que son relevantes en términos de expresión y de significancia estadística. A partir de esta lista filtrada se pueden crear gráficos de Heatmap, como el que se muestra en la figura 7, donde se ven aquellos genes que se expresan de diferente manera en la condición LL con respecto a la condición LP.

Las herramientas utilizadas en ese trabajo permiten a los usuarios biólogos ingresar datos y cambiar configuraciones a través de una sencilla interfaz gráfica que no requiere ningún tipo de conocimiento sobre el lenguaje de programación y los algoritmos que funcionan por debajo. Esto les permite abstraerse de la parte informática para centrarse en la interpretación biológica de los resultados.

Estos sistemas son muy útiles en el caso de proyectos multidisciplinarios que involucran a investigadores de diferentes lugares geográficos, debido a que permiten compartir datos manteniendo la seguridad y la privacidad de los mismos. Esto se debe en parte a su interfaz web que puede ser accedida desde cualquier lugar siempre que el usuario se autentifique de manera correcta. El proceso de autorización y de autenticación del usuario es llevado a cabo por la aplicación de seguridad CSM UPT, que permite manejar los usuarios de una manera centralizada para todas las aplicaciones del proyecto caBIG que se encuentren instaladas en el sistema.

El almacenamiento de datos en la aplicación caArray asegura el cumplimiento del estándar MIAME, por lo que pueden ser compartidos con otros investigadores o comparados con resultados de otros estudios. Este estándar creado por la *Functional Genomics Data (FGED) Society* es uno de los más utilizados en la actualidad por la comunidad científica.

Es importante señalar que cada laboratorio de biología tiene sus necesidades particulares que surgen como producto de las tecnologías que utilizan y de las hipótesis que plantean. Por ello es necesario que las herramientas utilizadas tengan la flexibilidad suficiente como para poder agregar las funcionalidades que sean necesarias. En este sentido, GenePattern permite crear nuevos módulos y rutinas que amplían las funciones estándar incluidas en la plataforma. En el presente trabajo se mostró cómo es posible adaptar una función escrita en R para su utilización dentro de la aplicación. Esto representa una característica muy importante ya que permite reutilizar código que ya ha sido ampliamente probado y mejorado en lugar de programar desde cero las funciones. Además se creó una rutina que permite automatizar el análisis llevado a cabo. Dicha rutina puede ser aplicada en otros experimentos que involucren el modelo de microarreglo Gene 1.0 ST de Affymetrix. Lo interesante de esto es que se trata de uno de los últimos modelos de la empresa y aún no existen muchos desarrollos al respecto.

111.Gonzalez.2010.NuGO  [Edit Sharing...](#)  [Show Input Parameters](#) 

-  **NuGOExpressionFileCreator1.input.filename:** parkinson.zip
-  **MADMAXArrayQualityAnalysis2.input.filename:** parkinson.zip56390zip

↳ step 1: MADMAXArrayQualityAnalysis [id: 113]  [Show Input Parameters](#) 

-  **input.filename:** file:/home/genepattern/Tomcat/webap...results/111/parkinson.zip56390zip
-  **AffyArrayQualityAnalysis_Results.zip** (5.8 MB)  (Last modified: Sun Oct 24 16:30:25 GMT+01:00 2010)
-  **stdout.txt** (48.0 KB)  (Last modified: Sun Oct 24 16:30:25 GMT+01:00 2010)

↳ step 2: NuGOExpressionFileCreator [id: 112]  [Show Input Parameters](#) 

-  **input.filename:** file:/home/genepattern/Tomcat/webapps/gp/jobResults/111/parkinson.zip
-  **6LA2.CEL** (10.6 MB)  (Last modified: Thu Jul 02 13:28:58 GMT+01:00 2009)
-  **4LP1.CEL** (10.6 MB)  (Last modified: Thu Jul 02 13:38:08 GMT+01:00 2009)
-  **3LL1.CEL** (10.6 MB)  (Last modified: Thu Jul 02 13:50:10 GMT+01:00 2009)
-  **12LP3.CEL** (10.6 MB)  (Last modified: Thu Jul 02 13:59:06 GMT+01:00 2009)
-  **11LL3.CEL** (10.6 MB)  (Last modified: Thu Jul 02 14:42:42 GMT+01:00 2009)
-  **2LA1.CEL** (10.6 MB)  (Last modified: Thu Jul 02 14:52:02 GMT+01:00 2009)
-  **1NA1.CEL** (10.6 MB)  (Last modified: Thu Jul 02 15:01:32 GMT+01:00 2009)
-  **5NA2.CEL** (10.6 MB)  (Last modified: Thu Jul 02 15:11:14 GMT+01:00 2009)
-  **10LA3.CEL** (10.6 MB)  (Last modified: Thu Jul 02 16:00:26 GMT+01:00 2009)
-  **9NA3.CEL** (10.6 MB)  (Last modified: Thu Jul 02 16:11:22 GMT+01:00 2009)
-  **7LL2.CEL** (10.6 MB)  (Last modified: Thu Jul 02 16:22:22 GMT+01:00 2009)
-  **8LP2.CEL** (10.6 MB)  (Last modified: Thu Jul 02 16:35:36 GMT+01:00 2009)
-  **ragene10stv1rmentrezg.chip** (1006.0 KB)  (Last modified: Sun Oct 24 16:04:51 GMT+01:00 2010)
-  **parkinson.gct** (3.9 MB)  (Last modified: Sun Oct 24 16:05:33 GMT+01:00 2010)
-  **stdout.txt** (30.0 KB)  (Last modified: Sun Oct 24 16:05:33 GMT+01:00 2010)

↳ step 3: LimmaAnalysis [id: 114]  [Show Input Parameters](#) 

-  **Normalized.Data:** parkinson.gct
-  **Target.File:** parkinson.cls
-  **log2.gct** (1.3 MB)  (Last modified: Sun Oct 24 16:30:39 GMT+01:00 2010)
-  **group1-group0-limmaResults.txt** (3.6 MB)  (Last modified: Sun Oct 24 16:30:49 GMT+01:00 2010)
-  **group2-group0-limmaResults.txt** (3.6 MB)  (Last modified: Sun Oct 24 16:30:53 GMT+01:00 2010)
-  **group3-group0-limmaResults.txt** (3.6 MB)  (Last modified: Sun Oct 24 16:30:56 GMT+01:00 2010)
-  **group2-group1-limmaResults.txt** (3.6 MB)  (Last modified: Sun Oct 24 16:30:59 GMT+01:00 2010)
-  **group3-group1-limmaResults.txt** (3.6 MB)  (Last modified: Sun Oct 24 16:31:03 GMT+01:00 2010)
-  **group3-group2-limmaResults.txt** (3.6 MB)  (Last modified: Sun Oct 24 16:31:06 GMT+01:00 2010)
-  **stdout.txt** (7.0 KB)  (Last modified: Sun Oct 24 16:31:06 GMT+01:00 2010)

Figura 6. Resultados del análisis utilizando la rutina

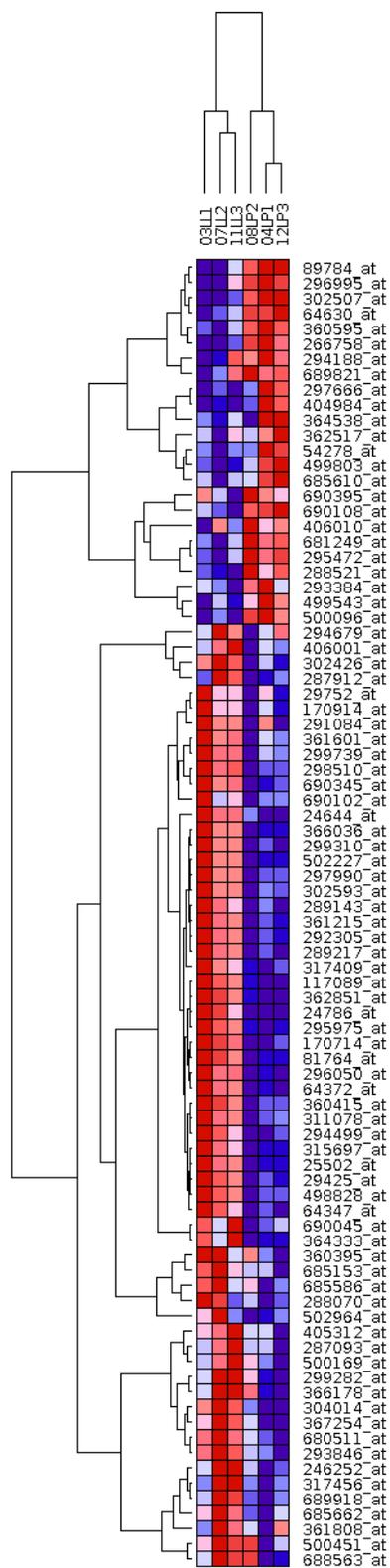


Figura 7. Heatmap para la comparación LL-LP

Estas herramientas fueron aplicadas sobre los datos de expresión generados por el Laboratorio de Parkinsonismo Experimental. A partir de este análisis se pudo encontrar un conjunto de genes que se expresan de diferente manera dependiendo del fármaco que se suministre. Además la utilización del CDF de MNBI permitió encontrar 48 genes diferenciales en la comparación LL-LP que no aparecen cuando se utiliza el CDF estándar. Esto se debe a que utiliza una anotación más detallada que la estándar. A partir de esta lista de genes, los biólogos seleccionaron un conjunto más pequeño teniendo en cuenta la información funcional y experimental existente. Estos genes fueron validados de forma experimental utilizando la PCR de tiempo real. Estos datos no son consignados en este trabajo.

Discusión

Si bien esta herramienta facilita en gran medida la tarea del biólogo, es importante resaltar que la utilización de algoritmos sin conocer completamente su funcionamiento puede llevar a resultados erróneos. Por eso estas herramientas deberían utilizarse como un primer acercamiento a los datos o bajo la supervisión de personal capacitado.

Agradecimientos

Este trabajo fue realizado gracias a los aportes realizados por la Agencia Nacional (Proyecto PICT-2008-0807), Agencia Córdoba Ciencia (PID-2008) y la UCC.

Referencias

- [1] Eric Tavela. caIntegrator2: Translational Web Portals On-Demand. caBIG Annual Meeting 2009
- [2] Gavai AK, de Groot PJ, Lin K, Hooiveld G, Liu Y, Nijveen H, Neerinx P, Müller M, Leunissen JAM (2008) MADMAX—Management and analysis database for multiplatform microArray experiments. BMC Bioinformatics (submitted). [https://madmax.bioinformatics.nl/pls/apex/wwv_flow_file_mgr.get_file?p_security_group_id=723402266026950&p_fname=About_MADMAX.pdf]
- [3] Lu X, Zhang X. The effect of GeneChip gene definitions on the microarray study of cancers. 2006. BioEssays 28(7) 739-746
- [4] Sandberg R, Larsson O. Improved precision and accuracy for microarrays using updated probe set definitions. 2007. BMC Bioinformatics 8:48
- [5] Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acid Research 33 (20), e175
- [6] Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003 Apr;4(2):249-64.
- [7] De Groot PJ, Reiff C, Mayer C, Müller M. NuGO contributions to GenePattern. Genes Nutr. 2008 December; 3(3-4): 143–146.

Datos de Contacto:

Germán González
Grupo de Minería de Datos en Biociencias
Av. Armada Argentina 3555, Córdoba, Argentina (CP 5017)
E-mail: german@bioinformaticos.com.ar.