

ENGINYERIA INFORMÀTICA

**2658-1: BIOINFORMÀTICA: INTERFAZ WEB PARA MOSTRAR LAS RELACIONES DE EXPRESIÓN
ENTRE GRUPOS DE GENES**

Memoria del Proyecto Final de Carrera
de Ingeniería Informática
realizado por
José Luis Aylas Flores
y dirigido por
Jordi González Sabaté
y codirigido por
Mario Huerta
Bellaterra, 17 de Setiembre de 2012

El sotasignat, Jordi Gonzàlez i Sabaté
Professor/a de l'Escola Tècnica Superior d'Enginyeria de la UAB

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat sota la seva direcció per en Jose Luis Aylas Flores.

I per tal que consti firma la present.

Signat:

Bellaterra, 17 de Setembre de 2012

El sotasignat, Mario Huerta
de l'empresa, Institut de Biotecnologia i de Biomedicina de la UAB

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat en
l'empresa sota la seva supervisió mitjançant conveni amb la Universitat
Autònoma de Barcelona.

Així mateix, l'empresa en té coneixement i dóna el vist-i-plau al contingut
que es detalla en aquesta memòria.

Signat:

Bellaterra, 17 de Setembre de 2012

Índice de contenido

1.- Introducción.....	5
1.1.- Motivación del proyecto.....	5
1.2.- Estado del arte.....	5
1.3.- Objetivos.....	9
1.4.- Organización de la memoria.....	10
2.- Fundamentos teóricos.....	11
2.1.- Biología y Bioinformática.....	11
2.2.- Microarrays.....	12
2.2.1.- Genes Marcadores.....	13
3.- Fases.....	14
3.1.- Análisis de Memorias/Directorios/Ficheros.....	14
3.1.1.- Adquisición de conocimientos.....	14
3.1.2.- Estructura del servidor.....	15
3.2.-Análisis del funcionamiento del preproceso.....	15
3.3.- Desarrollo de los nuevos cálculos del preproceso.....	16
3.4 Diseño y desarrollo de la interfaz web.....	16
4.- Informe Técnico.....	23
4.1 Estructura del servidor.....	23
4.1.1 Estructura de directorios.....	23
4.1.1.1 Representación de la arquitectura de directorios para el preproceso.....	24
4.1.1.1.1 Ficheros de entrada utilizados como entrada para los nuevos cálculos del preproceso.....	24
4.1.1.1.2 Programas desarrollados y los ficheros de salida generados en los nuevos cálculos del preproceso.....	25
4.1.1.2 Representación de la arquitectura de directorios para la aplicación web.....	29
4.1.1.2.1 PHPs desarrollados para la Aplicación web.....	29
5.- Conclusiones.....	30
6.- Referencias.....	32

NOMENCLATURA A TENER EN CUENTA

Cambio fenotípico: Es la dependencia existente entre las expresiones de 2 genes. Las expresiones de los genes no son arbitrarias y están todas interrelacionadas.

Relaciones de expresión: Es la dependencia existente entre las expresiones de 2 genes. Las expresiones de los genes no son arbitrarias y están todas relacionadas.

Relación de expresión no lineal: Es la expresión entre 2 genes pero que no siguen una relación de coexpresión o de inhibición. Es decir que no se expresan a la vez ($y=mx$) o que la expresión de un gen no significa que el otro gen deja de expresarse ($y=-mx$).

Coexpresión de genes: Genes que se expresan simultáneamente manteniendo una relación $y=mx$ en sus expresiones. Los genes que se expresan simultáneamente llevan a cabo un cambio fenotípico cuando se expresan.

Microarray: Tecnología que permite obtener los niveles de expresión de un gran número de genes (1500-10000) para un gran número de condiciones experimentales (100-1000).

1.-Introducción

1.1.-Motivación del Proyecto

Este ha sido trabajo que se encuentra englobado en un proyecto mayor que se esta realizando en el IBB (Institut de Biotecnología y de Biomedicina de la UAB) para el análisis de datos de microarrays.

Mi motivación inicial de este proyecto fue profundizar un poco más en lo que respecta a la Bioinformática. Si bien en un inicio la considere como una asignatura optativa, durante el proceso del desarrollo de mi proyecto en el campo de la Bioinformática he logrado entender la orientación, usos y objetivos de la misma.

Mi proyecto de final de carrera ha representado una gran oportunidad para poner a prueba los conocimientos y habilidades que he adquirido durante todo el trascurso de mi formación universitaria. El saber que el proyecto estaba enmarcado en un ambiente de investigación interdisciplinar que relaciona la informática con la biología y la medicina lo hacía también muy interesante.

En general, el desarrollo de mi proyecto me ha permitido afianzar mi formación como informático y además me ha aportado formación en el campo de la biotecnología. He afianzado mis conocimientos informáticos porque gracias a la realización de este proyecto he podido aplicar conocimientos teóricos adquiridos durante la carrera a un problema real, que muchas veces creemos solo serán utilizados para aprobar exámenes y que he puesto en práctica tanto durante el preproceso de cálculo como para el desarrollo de la interfaz web.

1.2 Estado del arte

La BioInformática es la ciencia que estudia los fenómenos biológicos (en nuestro caso de la biología molecular) desde un punto de vista computacional. Su finalidad es poder ofrecer métodos robustos para la comprensión, simulación y predicción de los comportamientos observados en los seres vivos. Por consiguiente las principales aplicaciones de la Bioinformática consisten en la simulación, minería de datos

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

(DataMining), análisis de la expresión génica, predicción de estructuras de proteínas, estudio de las interacciones proteína – proteína, entre otros.

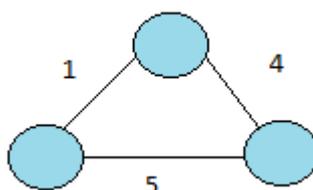
Los genes al expresarse sintetizan las diferentes proteínas y estas proteínas, una vez sintetizadas, son las que llevan a cabo las diferentes funciones de la célula. De esta forma los genes, al expresarse, determinan el estado celular y modificando su expresión, provocan cambios celulares. Estos cambios celulares pueden significar el paso de la enfermedad a la salud o todo lo contrario y es por ello, que el estudio de los genes y sus relaciones de expresión es tan importante. El estudio de los cambios celulares, y la relación que mantienen los diferentes procesos celulares entre ellos (activación, desactivación,...) es uno de los principales problemas a resolver por parte de la biotecnología, la bioinformática y la medicina molecular.

La tecnología de microarray permite obtener los niveles de expresión de un gran número de genes (1500-10000) para un gran número de condiciones experimentales (100-1000). En el IBB se está desarrollando un servidor para el análisis de microarrays [8][9][10][11][12] para poder estudiar entre otros, las relaciones de expresión entre genes.

Los genes cuyas expresiones mantienen una relación lineal dan lugar a un determinado proceso biológico. Los genes cuyas expresiones mantienen una relación no lineal, relacionan procesos diferentes. Todo esto podemos conocerlo analizando los datos de microarray, de forma que podemos realizar el estudio de la relación entre procesos estudiando las relaciones de expresión no lineales entre genes que participan en los procesos que se quieren relacionar.

Para estudiar esta relación entre las expresiones de los genes se utiliza un modelo abstracto de datos: Los Grafos. Más concretamente, los grafos cliques. Un clique es un grafo tal que para cada par de nodos (genes) del grafo existe una arista que los conecta [6]. Las aristas de los cliques representan la relación de expresión no lineal entre cada nodo (gen) del clique. Cada gen del clique pertenece a un proceso diferente puesto que si perteneciesen al mismo proceso sus expresiones mantendrían una relación lineal, tal como hemos comentado anteriormente. De esta forma, a partir de la relación no lineal entre las expresiones de los genes de dos nodos se puede inducir la relación que mantienen los procesos a los que pertenecen cada uno de estos genes.

Para encontrar los cliques disponemos de la aplicación MACE (Maximal Clique Enumerator) [7], además de otros múltiples módulos desarrollados en el IBB. Estos módulos están disponibles en el servidor para el análisis de microarrays: <http://revolutionresearch.uab.es> [10] que nos permite obtener los cliques clasificados según los tipos de relaciones no lineales existentes entre los genes del clique. Esta clasificación por la tipología de las relaciones no lineales es importante porque cada tipología implica un tipo de relación entre los genes diferente (activación, desactivación, etc....).



Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

Figura 1.1: Clique de 3 nodos. Los nodos representan a los genes de diferentes procesos y las aristas representan la relación de expresión no lineal que hay entre estos genes. Estas relaciones no lineales tienen diferentes tipologías y cada tipología tendrá un significado biológico diferente.

Dos cliques serán isomorficos lineales entre sí, sí y solo si además de representar sus aristas las mismas tipologías, cada gen de uno de los cliques mantiene una relación de expresión lineal con un gen diferente del otro clique. Estos genes cuyas expresiones mantienen una relación lineal pertenecerán al mismo proceso y su relación no lineal con las expresiones de los otros pares de genes describirá la relación con el resto de procesos.

Los cliques isomorficos lineales son un grupo de cliques isomorficos lineales respecto a un clique dado.

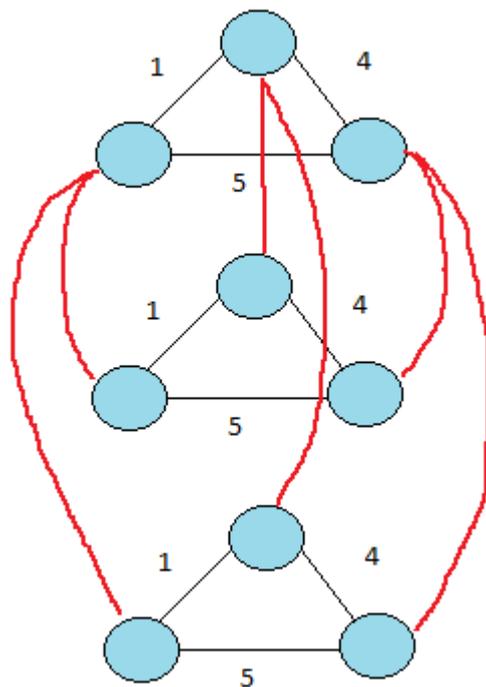


Figura 1.2: grupo de Cliques Isomorficos Lineales respecto a un clique dado. Todos los Cliques tienen la misma tipología en las aristas y son cliques lineales respecto a el primer clique.

En el servidor del IBB [10] están disponibles los cliques hallados mediante la aplicación MACE. A partir de estos cliques han sido encontrados mediante los módulos desarrollados en el IBB los grupos de cliques isomorficos lineales.

El paso lógico siguiente es encontrar cliques isomorficos lineales en los que los cliques sean lineales todos con todos, a los que llamaremos clique de cliques. Estos cliques de cliques realmente nos describirán la relación entre procesos, puesto que no se limitarán a la relación entre dos genes pertenecientes a diferentes

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

procesos, sino que abarcarán todo el conjunto de genes de cada proceso que mantienen una determinada relación de expresión no lineal con los genes del otro proceso.

1.3 Objetivos

El preproceso son los cálculos que generan los ficheros que se utilizaran en la aplicación web del servidor <http://revolutionresearch.uab.es> [10]. El preproceso no es interactivo y se ejecuta una sola vez para cada microarray que se ha de analizar.

La Interfaz web son el conjunto de elementos que permiten al usuario realizar acciones sobre el servidor para el análisis de microarrays <http://revolutionresearch.uab.es> [10]. Por lo mismo, se considera parte de la interfaz a sus elementos de identificación, de navegación, de contenidos y de acción.

Los objetivos a alcanzar son:

1. En el preproceso:
 - a. Obtener a partir de los *cliques isomorficos lineales* los *cliques de cliques*.
 - i. Convertir el formato de los ficheros que contienen los *grupos de cliques isomorfos lineales* a un formato adecuado para el *MACE*.
 - b. Clasificación de los *cliques de cliques*.
 - i. Clasificación de los *cliques de cliques* por la tipología de sus aristas.
 - ii. Convertir el formato de los ficheros de *cliques de cliques* a un formato en que cada clique este representado por los genes que conforman los cliques.
 - c. Preparar los datos, ficheros y carpetas para que la interfaz web tenga que realizar los menos cálculos posibles.
2. Desarrollar una interfaz web donde:
 - a. Se muestren los *cliques de cliques* donde:
 - i. Se muestren gráficamente las diferentes tipologías para las que se han encontrado *cliques de cliques*.
 - ii. Se muestren estadísticas para cada tipología como el número de *cliques de cliques* hallados, el número de cliques totales, etc.
 - iii. Se muestren los *cliques de cliques* pertenecientes a cada tipología.
 - iv. Se muestre en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario.

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

- b. Se muestren los *grupos de cliques isomorficos lineales* donde:
 - i. Se muestren gráficamente las tipologías para las que se han encontrado *grupos de cliques isomorficos lineales*.
 - ii. Se muestren estadísticas para cada tipología, como el número de *grupos de cliques isomorficos lineales* hallados, el número de cliques totales, etc.
 - iii. Se muestren los *grupos de cliques isomorficos lineales* pertenecientes a cada tipología.
 - iv. Se muestre en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario en detalle.
- c. Se muestren los cliques donde:
 - i. Se consideran solo los cliques que tienen un mínimo del 35% de relaciones no lineales que no son de baja curvatura.
 - ii. Se muestran gráficamente las tipologías para las que se han encontrado cliques que han superado el filtro.
 - iii. Se muestran Los cliques pertenecientes a cada tipología con su factor de correlación.
 - iv. Se muestra en una nueva ventana la relación de expresión en detalle entre los genes de los cliques elegidos por el usuario.

1.4. Organización de la memoria

Para alcanzar los objetivos anteriormente descritos se ha seguido la planificación que se expone a continuación.

El trabajo realizado ha tenido una duración de 10 meses. Se inició en noviembre de 2011 y finalizó en agosto de 2012. Este trabajo se divide en tres procesos bien diferenciados:

- a) Análisis de memorias, directorios y ficheros del servidor para análisis de microarrays: <http://revolutionresearch.uab.es> [10]
- b) Desarrollo del preproceso a partir de los datos previamente generados.
- c) Diseño y desarrollo de la interfaz web.

En la siguiente sección de la memoria, expongo los fundamentos teóricos necesarios para comprender los conceptos con los que he trabajado a lo largo del proyecto.

Acto seguido detallo las fases que forman el diseño y la implementación del preproceso y de la interfaz web. Primero me centro en el desarrollo del algoritmo diseñado para hallar los cliques de cliques a partir de los cliques isomorfos lineales. Finalmente termino con la interfaz web desarrollada, explicando cada uno de sus elementos.

La estructura seguida por la memoria y su contenido sería el siguiente:

- Fundamentos teóricos: Donde expongo los conocimientos biológicos y técnicos necesarios para comprender el proyecto.
- Fases: Donde describo el trabajo desarrollado para alcanzar los objetivos propuestos, los problemas encontrados y las soluciones adoptadas.
- Informe técnico: Describo los programas implementados y la estructura de los ficheros y directorios del servidor.
- Conclusiones: Expongo las conclusiones del proyecto

2. Fundamentos teóricos

2.1 Biología y BioInformática

La información biológica se encuentra codificada en los genes y se expresa mediante los genes. Esta idea se refleja en el Dogma Central de la Biología Molecular, es decir que el ADN es transcrito a ARN mensajero y que éste es traducido a proteína, elemento que finalmente realiza la acción celular.

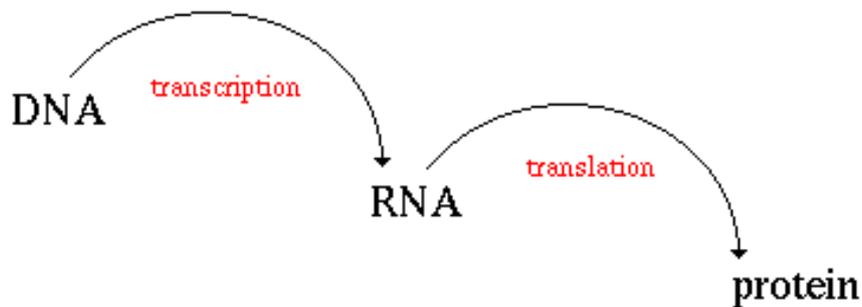


Figura 1.4.1 Dogma Central de la Biología

La biología se enfrenta con el problema de la decodificación del lenguaje biológico, el saber como se codifica la información en los genes, como y cuando se traduce esta información, el que determina la estructura de las proteínas, como se determina la función de las proteínas, así pues la bioinformática sirve para estudiar como se procesa toda esta información biológica.

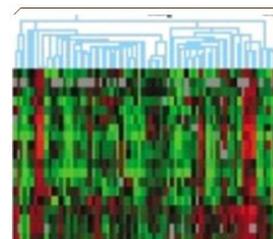
La Bioinformática es la disciplina científica que engloba todos los aspectos de la adquisición, procesamiento, distribución, análisis, interpretación e integración de la información biológica.

```
ATGTGCAATGCTT
CGTTACGGCTCAA
TATGCCGCAGTAA
GCTGCAGTATCCG
CCGCAGTAACTGG
GCCGCAG.....
```

Datos



Recursos



Conocimiento

2.2 Microarray

La tecnología de microarray hace uso de los recursos de secuencias creados por los proyectos del genoma y otros esfuerzos de secuenciación para decir que genes se expresan en un tipo celular particular de un organismo, en un momento determinado y en condiciones determinadas. Por ejemplo, permitir la comparación de la expresión de genes entre normales y enfermos.

En la era pre-genómica la biología trataba a los genes individualmente, de uno en uno, así cada gen se estudiaba a fondo. En la era post-genómica se pueden estudiar muchos genes a la vez.

En resumen, la tecnología de microarrays se basa en someter una gran cantidad de genes a muchas condiciones diferentes, lo que permite obtener su nivel de expresión. Así para cada una de estas condiciones, que llamare condiciones muestrales, se obtiene el nivel de expresión de cada gen. Es decir, en la tecnología de microarrays, cada una de las k características va a corresponder a cada una de las condiciones muestrales a las que se someten cada uno de los n genes.

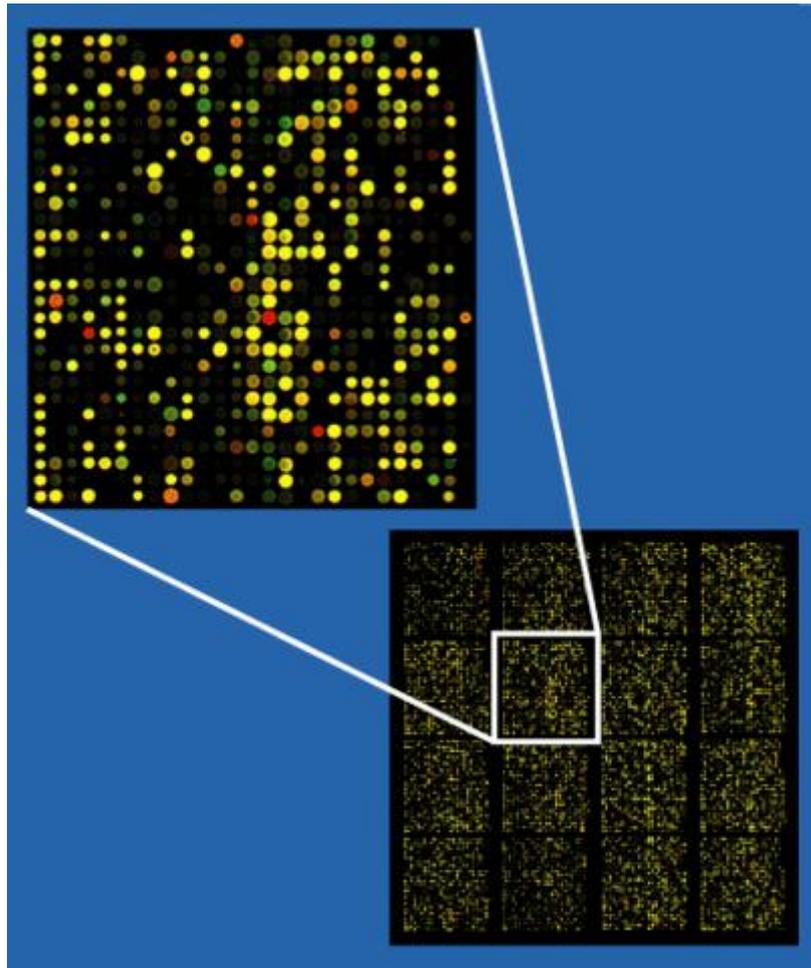


Figura x.x Microarray en detalle

La tecnología microarray muestra la actividad de miles de genes simultáneamente. Esta capacidad de alto rendimiento es ideal para la era post-genómica en que los proyectos de secuenciación están produciendo grandes cantidades de datos.

La tecnología actual de microarrays me permite evaluar alrededor de 10000 genes bajo un número de condiciones muestrales de alrededor de 1000. Para el caso de la microarray 17 del servidor del IBB, que es con la que he realizado las pruebas, se contemplan 1416 genes(individuos) sometidos a 118 condiciones muestrales (características). Es decir, obtengo un total de 167.088 evaluaciones de nivel expresión génica.

2.3 Genes Marcadores

Un gen marcador es aquel gen que con su expresión está marcando la presencia de un proceso, condición celular, o patología concreta. En análisis de microarrays, en concreto, un gen marcador es un gen que se sobreexpresa en determinadas condiciones experimentales de la microarray y se infraexpresa en otras. Con lo que un gen marcador, para un ejemplo concreto de tejidos sano y tejido tumoral, el gen marcador marca la presencia de su condición celular, y marca entonces si el tejido es sano o tumoral.

3.-Fases

A continuación se muestra una tabla comparativa entre la planificación inicial y la planificación final.

Planificación Inicial	Planificación Final
Análisis de Memorias/Directorios/Ficheros para seguir con el desarrollo de los nuevos cálculos del preproceso.	Análisis de Memorias/Directorios/Ficheros
Análisis del funcionamiento del preproceso para comprender como continuar los nuevos cálculos del preproceso. Desarrollo del preproceso siguiendo con la siguiente fase	Análisis del funcionamiento del Preproceso Desarrollo del preproceso
Diseño y desarrollo de la interfaz web.	Diseño y desarrollo de la interfaz web Estadísticas añadidas para cada interfaz web realizada
Memoria	Memoria

Para las pruebas se ha usado una microarray proporcionada por la base de datos del Instituto Nacional del Cáncer (EE.UU.) (Scherf et al 2000)[35] . Corresponde a los perfiles de 9.703 cDNAs que representa ~ 8000 genes únicos de 60 líneas celulares, en relación con los perfiles de actividad de 1400 medicamentos (finalmente resumidos en 1417 genes de 40 líneas celulares, en relación a 118 medicamentos, que representan los valores más significativos). A esta microarray se la identifica como m17.

A continuación en los siguientes apartados se expone la planificación de las distintas fases, los problemas que han surgido y como se han solucionado con éxito.

3.1 Análisis de Memorias/Directorios/Ficheros

3.1.1 Adquisición de conocimientos

Para realizar correctamente el trabajo que conllevaba el proyecto era necesario aprender conceptos biológicos fundamentales para poder entender el proyecto que se ha desarrollado.

Los conceptos adquiridos fueron pues referentes al análisis de microarrays y la gran cantidad de datos que manejan y como la información biológica se encuentra codificada en los genes y se expresa mediante los genes.

Para realizar la implementación se ha tenido que aprender a utilizar la herramienta web PCOPGene[14] leyendo manuales de ayuda de dicha herramienta.

Tal y como se ha visto para estudiar esta relación entre las expresiones de los genes se ha utilizado un modelo abstracto de datos: Los Grafos. Más concretamente, los grafos de cliques. Un clique es un grafo tal que para cada par de nodos (genes) del grafo existe una arista que los conecta [6].

3.1.2 Estructura del servidor

Para el desarrollo de la aplicación tuve que partir del análisis de los datos iniciales ya disponibles por lo cual tuve que guiarme de la memoria del anterior trabajo, para tener una referencia de hasta donde se había llegado y poder partir desde ese punto.

Tuve que hacer el análisis de la estructura del servidor para saber que ficheros estaban disponibles y en que carpetas se encontraban. Ya que se partía de una gran cantidad de directorios y ficheros era altamente recomendable hacer este análisis.

El saber de qué ficheros se disponían y entender el formato en el que estaban representados era clave para la siguiente fase.

3.2 Análisis del funcionamiento del preproceso:

Para continuar con el desarrollo del preproceso tuve que entender lo que hacía hasta ese momento el preproceso. Con lo cual, tuve que analizar los pasos que seguía, es decir, de que ficheros necesitaba y que ficheros generaba.

Para el análisis pues del preproceso, analice el fichero que se encargaba de ejecutar el preproceso.

En el fichero encargado de ejecutar el preproceso se ejecutan : fase1, fase2 y fase3

Por consiguiente tenía que analizar en que consistía cada fase.

Fase1: Esta fase incluía una serie de ficheros a ejecutarse en un determinado orden. Lo que generan son los cliques.

Fase2: En esta fase que incluía una serie de ficheros a ejecutarse en un determinado orden. Lo que generan son los cliques isomorfos lineales y no lineales.

Fase3: En esta fase que incluía una serie de ficheros a ejecutarse en un determinado orden. Lo que generan son los grupos de cliques isomorfos lineales.

Por lo consiguiente hasta la fase 3 del preproceso se disponía de los grupos de cliques isomorfos lineales. Se quería pues seguir con el preproceso con una siguiente fase, la fase 4, que mostraría los cliques de los grupos de cliques isomorfos lineales, es decir, los cliques de cliques.

3.3 Desarrollo de los nuevos cálculos del preproceso

Se disponía del preproceso hasta la fase 3, se quería pues añadir una cuarta fase.

En la cuarta fase pues se pretendía hallar los cliques de los grupos de cliques isomorficos lineales, es decir , los cliques de cliques.

Partía pues de los grupos de cliques isomorficos lineales y contaba con el programa MACE, con el que podría hallar cliques. Lógicamente tenía que pasarle al MACE los grupos de cliques isomorficos lineales para poder hallar sus cliques, y obtener los cliques de cliques.

Había que obtener a partir de los grupos de cliques isomorficos lineales los cliques de cliques, pero se partía que los grupos de cliques isomorficos lineales no tenían un formato aceptado por el MACE, así que se decidí convertir el fichero de los grupos de cliques isomorficos lineales a un formato aceptado por el MACE. Con el fichero en un formato adecuado se generan los cliques de cliques sin clasificación por la tipología de sus tipos de aristas.

Luego había que tener una clasificación de los cliques de cliques por la tipología de sus aristas porque se requería poder acceder a estos cliques de cliques por los tipos de curva de los cliques así se decidió generar un fichero que tuviera el formato de los grupos de cliques isomorficos lineales siguiendo el formato del fichero de los grupos de cliques isomorficos lineales consiguiendo la clasificación de los cliques de cliques por la tipología de sus aristas. También se quería convertir el formato de los ficheros de cliques de cliques a un formato en que cada clique estuviera representado por los genes de los cliques que lo conformaban para poder presentarlos en una interfaz web donde se vieran los genes que representaban los cliques , por consiguiente se decidió generar un fichero en la que cada clique estuviera representado por los genes de los cliques que lo conformaban a partir de un fichero donde se encontraban todos los cliques mediante la representación de los genes que los conformaban.

El último paso para solventar los problemas surgidos en los nuevos cálculos del preproceso fue dejar los datos, ficheros y carpetas accesibles a la aplicación web.

3.4. Diseño y desarrollo de la interfaz web

Mi aplicación web consistía en el desarrollo de tres herramientas web:

- Una interfaz web para los cliques de cliques, con una vista general en que aparecen las tipologías y su correspondiente interfaz web en detalle, donde se muestran los cliques de cliques de cada tipología representados por sus genes.
- Una interfaz web para los grupos de cliques isomorficos lineales, con una vista general en que aparecen las tipologías y su correspondiente interfaz web en detalle, donde se muestran los grupos de cliques isomorficos lineales de cada tipología representados por sus genes.
- Una interfaz web para los cliques isomorficos lineales y no lineales, con una vista general en que aparecen las tipologías y su correspondiente interfaz web en detalle, donde se los cliques isomorficos lineales y no lineales de cada tipología representados por sus genes.

Se quería mostrar gráficamente las diferentes tipologías para las que se han encontrado cliques de cliques así que para conseguir estas gráficas recurrí fichero donde se encuentran los cliques de cliques y el fichero donde estaban todos los tipos de curva y así viendo para cada clique del

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

fichero el tipo de curva que le correspondía pude buscar el tipo de curva que le correspondía a cada clique.

Se quería que se mostrasen estadísticas para cada tipología como el número de cliques de cliques generados, el número de cliques totales, etc, así que para mostrar las estadísticas para cada tipología, tuve que contar el número de veces que aparecían los cliques y la cantidad de cliques de cliques.

Se requerían los cliques de cliques pertenecientes a cada tipología, así que recurrí al fichero donde se encuentran los cliques de cliques representados por los genes de los cliques que lo conforman.

A partir de estas páginas web desarrolladas se quería mostrar muestre en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario, así que mediante las opciones de poder elegir por checkboxes los genes que forman parte de los cliques pude mostrar en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario.

Se quería mostrar gráficamente las diferentes tipologías para las que se han encontrado los grupos de cliques isomorficos lineales así que para conseguir estas gráficas recurrí fichero donde se encuentran los grupos de cliques isomorficos lineales y el fichero donde estaban todos los tipos de curva y así viendo para cada clique del fichero el tipo de curva que le correspondía pude buscar el tipo de curva que le correspondía a cada clique.

Se quería que se mostrasen estadísticas para cada tipología como el número de grupos de cliques isomorficos lineales generados, el número de cliques totales, etc, así que para mostrar las estadísticas para cada tipología, tuve que contar el número de veces que aparecían los cliques y la cantidad de grupos de cliques isomorficos lineales.

Se requerían los grupos de cliques isomorficos lineales pertenecientes a cada tipología, así que recurrí al fichero donde se encuentran grupos de cliques isomorficos lineales representados por los genes de los cliques que lo conforman.

Y a partir de estas páginas web desarrolladas se quería mostrar muestre en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario, así que mediante las opciones de poder elegir por checkboxes los genes que forman parte de los cliques pude mostrar en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario.

Se quería mostrar gráficamente las diferentes tipologías para las que se han encontrado los cliques isomorficos lineales y no lineales y además solo considerar los cliques que tuvieran un minimo del 35% de relaciones no lineales que no fueran de baja curvatura, así que tuve que recurrir al fichero donde están todos los tipos de curvatura y analizar el tipo de curvatura de cada clique para poder garantizar que los cliques tuvieran así el mínimo del 35% de relaciones no lineales

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

Se requerían los cliques isomorficos lineales y no lineales pertenecientes a cada tipología, así que recurrí al fichero donde se encuentran cliques isomorficos lineales y no lineales representados por los genes de los cliques que lo conforman.

Y a partir de las páginas web desarrolladas se quería mostrar muestre en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario, así que mediante las opciones de poder elegir por checkboxes los genes que forman parte de los cliques pude mostrar en detalle en una nueva ventana la relación entre las expresiones de los genes elegidos por el usuario.

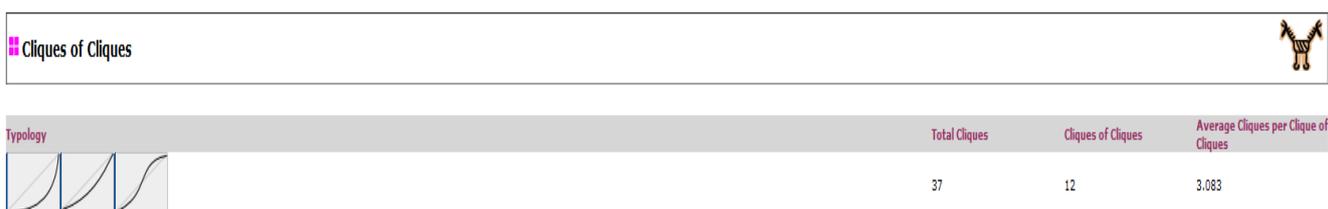


Figura 3.1 Vista de la interfaz web de los cliques de cliques. Se muestra la lista de las distintas tipologías de los clique de clique mediante las gráficas de los distintos tipos de curvas de las aristas de los cliques de los grupos de cliques de cliques. El número total de cliques es el número total de cliques que forman parte de un clique de cliques generados para la tipología mostrada. El número de cliques de cliques es el número de cliques de cliques generados para la tipología mostrada. El promedio de cliques de cliques por clique, es el promedio del total de cliques de cliques por el número total de cliques que forman parte de un clique de cliques.

En [figura 3.1] se muestra la lista de los cliques de cliques clasificados por la tipología de sus aristas, en la tipología se observan 3 tipos de curvas, la curva del tipo 0, la curva del tipo 1 y la curva del tipo 5, mostrando así las gráficas correspondiente a cada tipo de curva. Así para esta tipología de este clique de clique se muestra el número total de cliques que forman parte de un clique de cliques: 37, el número total de cliques de cliques es : 12 y el promedio del número total de cliques que forman parte de un clique de clique por el total de cliques de cliques: 3.083.

Id	Name	Id	Name	Id	Name
740	SCYA5	753	SID	763	FSTLL1
541	PDLIM7	714	DUSP1	696	FERMT2
538	STMN4	779	H.sapiens	718	ARHGGEF17
740	SCYA5	753	SID	763	FSTLL1
541	PDLIM7	714	DUSP1	696	FERMT2
538	STMN4	775	H.sapiens	718	ARHGGEF17
541	PDLIM7	745	ITGA3	648	RXRA
432	ANXB	679	PCSK5	716	C6orf108
724	TPM1	738	MS4A7	710	MT1X

Figura 3.2 Vista de la interfaz web de los cliques de cliques en detalle para una tipología dada. Cada línea representa un clique, en cada clique se muestran los genes que conforman el clique, cada grupo de cliques de cliques están separados por una línea rosa. En cada línea se tiene los genes que conforman el clique, con su identificador, el nombre del gen, genes más correlacionados con el gen seleccionado, y los microarrays para las cuales el gen seleccionado es un gen marcador. Mediante el botón de Interfaz

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

gráfica se pueden seleccionar como mínimo 2 genes y hasta un máximo de 7 genes para poder visualizar en detalle la relación entre las expresiones de los genes elegidos del usuario.

En [figura 3.2] se muestran los cliques de cliques representados por los genes que conforman los cliques para un tipología concreta. Cada línea representa un clique, en cada clique se muestran los genes que conforman el clique. Cada agrupación de cliques está separada por una línea rosa.

A cada gen le acompaña:

Su identificador: este identificador es un identificador único que se le asigna a cada gen para la microarray analizada.

Su nombre: el nombre del gen es el nombre con el que se le conoce en el Centro Nacional para la Información Biotecnológica (NCBI).

- Sus genes correlacionados: Genes cuyas expresiones tienen una mayor correlación con el gen dado.
- Genes marcadores: Muestra los microarrays para los cuales el gen seleccionado es un gen marcador.

Así el primer grupo de clique de cliques está formado por 3 cliques (una línea por clique):

En la primera línea se muestra un clique representado por los genes 740, 753 y 763.

En la segunda línea se muestra un clique representado por los genes 541, 714, 696.

En la tercera línea se muestra un clique representado por los genes 538, 779, 718.

Para visualizar en detalle, en una interfaz gráfica, la relación entre expresiones de genes elegidas por el usuario disponemos de las cajas de selección (checkboxes) al lado de cada gen, se ha de seleccionar entonces como mínimo 2 genes y se permite un máximo de hasta 7 genes.



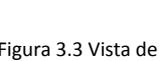
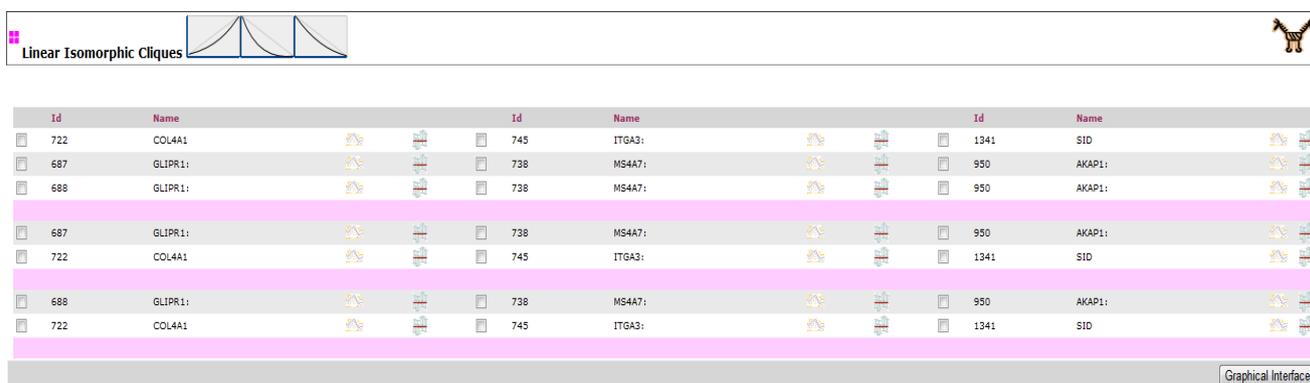
Typology	Total Cliques	Total Groups of Linear Isomorphic Cliques	Average Cliques per Group of Linear Isomorphic Cliques
	7	3	2.333
	4	2	2.000
	290	70	4.143
	8	4	2.000
	18	8	2.250
	32	12	2.667
	8	4	2.000

Figura 3.3 Vista de la interfaz web para los grupos de cliques isomorficos lineales. Se muestra la lista de las distintas tipologías de los grupos de cliques isomorficos lineales mediante las gráficas de los distintos tipos de curvas de las aristas de los cliques de los grupos de cliques isomorficos lineales. El número total de cliques es el número total de cliques que forman parte de un grupo de cliques isomorficos lineales generados para la tipología mostrada. El número total de grupos de cliques isomorficos lineales es el número de grupos de cliques isomorficos lineales generados para la tipología mostrada. El promedio de grupos de cliques isomorficos

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

lineales por cliques, es el promedio del total de grupos isomorfos lineales por el número total de cliques. que forman parte de un grupo de cliques isomorfos lineales.

En [figura 3.3] se muestran los grupos de cliques isomorfos lineales respecto a un clique dado clasificados por la tipología de sus aristas. En la primera tipología se observan 3 tipos de curvas, la curva del tipo 1, la curva del tipo 9 y la curva del tipo 10, mostrando así las gráficas correspondiente a cada tipo de curva. Así para esta tipología de este grupo de cliques isomorfos lineales se muestra el número total de cliques que forman parte de un grupo de cliques isomorfos lineales: 7, el número total de grupos de cliques isomorfos lineales: 12 y el promedio del total de grupos de cliques isomorfos lineales por el número total de cliques que forman parte de un grupo de cliques isomorfos lineales: 3.083.



Id	Name	Id	Name	Id	Name
722	COL4A1	745	ITGA3:	1341	SID
687	GLIPR1:	738	MS4A7:	950	AKAP1:
688	GLIPR1:	738	MS4A7:	950	AKAP1:
687	GLIPR1:	738	MS4A7:	950	AKAP1:
722	COL4A1	745	ITGA3:	1341	SID
688	GLIPR1:	738	MS4A7:	950	AKAP1:
722	COL4A1	745	ITGA3:	1341	SID

Figura 3.4 Vista de la interfaz web para los grupos de cliques isomorfos lineales en detalle para una tipología dada. Cada línea representa un clique, en cada clique se muestran los genes que conforman el clique, cada grupo de cliques isomorfos lineales están separados por una línea rosa. En cada línea se tiene los genes que conforman el clique, con su identificador, el nombre del gen, genes más correlacionados con el gen seleccionado, y los microarrays para las cuales el gen seleccionado es un gen marcador. Mediante el botón de Interfaz gráfica se pueden seleccionar como mínimo 2 genes y hasta un máximo de 7 genes para poder visualizar en detalle la relación entre las expresiones de los genes elegidos del usuario.

En [figura 3.4] se muestran los grupos de cliques isomorfos lineales respecto a un clique dado representados por los genes que conforman los cliques. Cada línea representa un clique, en cada clique se muestran los genes que conforman el clique. Cada agrupación de cliques está separada por una línea rosa.

A cada gen le acompaña:

Identificador del gen: este identificador es un identificador único que se le asigna a cada gen para la microarray analizada.

Nombre del gen: el nombre del gen es el nombre con el que se le conoce en el Centro Nacional para la Información Biotecnológica (NCBI).

Genes correlacionados: Genes cuyas expresiones tienen una mayor correlación con el gen dado.

Genes marcadores: Muestra los microarrays para los cuales el gen seleccionado es un gen marcador.

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

Así el primer grupo de cliques isomorfos lineales está formado por 3 cliques (una línea por clique):

En la primera línea se muestra un clique representado por los genes 722, 745, 1341.

En la segunda línea se muestra un clique representado por los genes 687, 738, 950.

En la tercera línea se muestra un clique representado por los genes 688, 738, 950.

Para visualizar en detalle, en una interfaz grafica, la relación entre expresiones de genes elegidas por el usuario disponemos de las cajas de selección (checkboxes) al lado de cada gen, se ha de seleccionar entonces como mínimo 2 genes y se permite un máximo de hasta 7 genes.

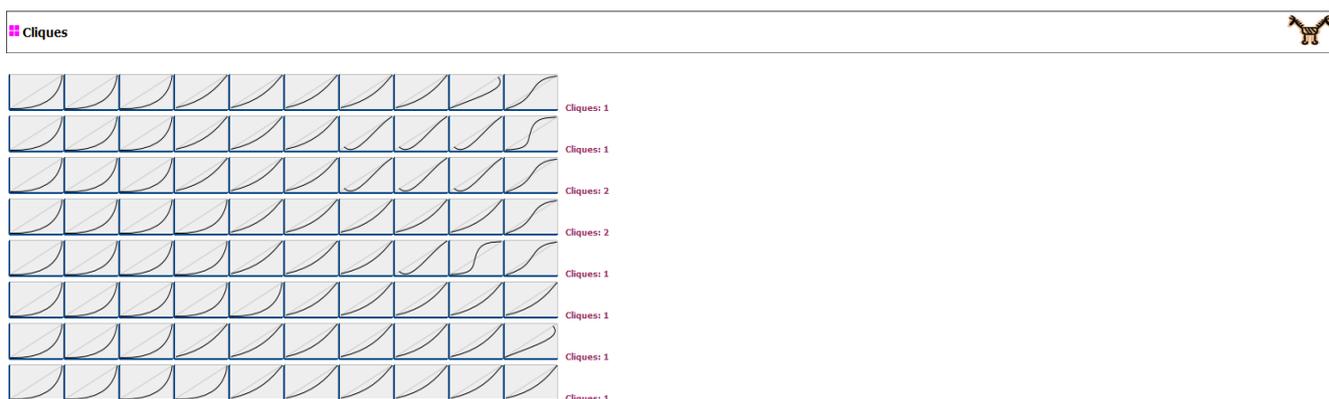


Figura 3.5 Vista de la interfaz web para los cliques. Se muestra la lista de las distintas tipologías de los de cliques isomorficos lineales y no lineales mediante las gráficas de los distintos tipos de curvas de las aristas de los cliques isomorficos lineales y no lineales que superen un coeficiente de correlación dado. El número de cliques es el número de cliques isomorfos lineales y no lineales que superan un coeficiente de correlación dado generados para la tipología mostrada.

En [figura 3.5] se muestran los cliques isomorfos lineales y no lineales clasificados por la tipología de sus aristas. En la primera tipología se observan 10 tipos de curvas, la curva del tipo 0, una segunda del tipo 0, una tercera del tipo 0, una del tipo 1, una segunda del tipo 1, una tercera del tipo 1, una cuarta del tipo 1, una quinta del tipo 1, una del tipo 2 y una del tipo 5, mostrando así las gráficas correspondiente a cada tipo de curva. Así para esta tipología de estos cliques isomorfos lineales y no lineales que superan un coeficiente de correlación dado el número de cliques es 1.

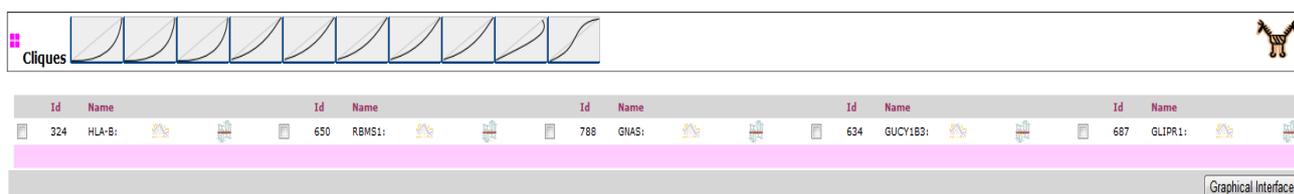


Figura 3.6 Vista de la interfaz web para los cliques isomorfos lineales y no lineales en detalle para una tipología dada que superan un coeficiente de correlación dado. Cada línea representa un clique, en cada clique se muestran los genes que conforman el clique, cada grupo de cliques isomorfos lineales están separados por una línea rosa. En cada línea se tiene los genes que conforman el clique, con su identificador, el nombre del gen, genes más correlacionados con el gen seleccionado, y los microarrays para las cuales el gen seleccionado es un gen marcador. Mediante el botón de Interfaz gráfica se pueden seleccionar como mínimo 2 genes y hasta un máximo de 7 genes para poder visualizar en detalle la relación entre las expresiones de los genes elegidos del usuario.

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

En [figura 3.6] se muestran los cliques isomorfos lineales y no lineales que superan un coeficiente de correlación dado representados por los genes que conforman los cliques. Cada línea representa un clique, en cada clique se muestran los genes que conforman el clique. Cada agrupación de cliques está separada por una línea rosa.

A cada gen le acompaña:

- a. Su identificador: este identificador es un identificador único que se le asigna a cada gen para la microarray analizada.
- b. Su nombre: el nombre del gen es el nombre con el que se le conoce en el Centro Nacional para la Información Biotecnológica (NCBI).
- c. Sus genes correlacionados: Genes cuyas expresiones tienen una mayor correlación con el gen dado.
- d. Genes marcadores: Muestra los microarrays para los cuales el gen seleccionado es un gen marcador.

Así el primer clique isomorfo lineal o no lineal (una línea por clique):

En la primera línea se muestra un clique isomorfo lineal o no lineal representado por genes cuyos identificadores son 324, 650, 788, 634, 687.

Para visualizar en detalle, en una interfaz gráfica, la relación entre expresiones de genes elegidas por el usuario disponemos de las cajas de selección (checkboxes) al lado de cada gen, se ha de seleccionar entonces como mínimo 2 genes y se permite un máximo de hasta 7 genes.

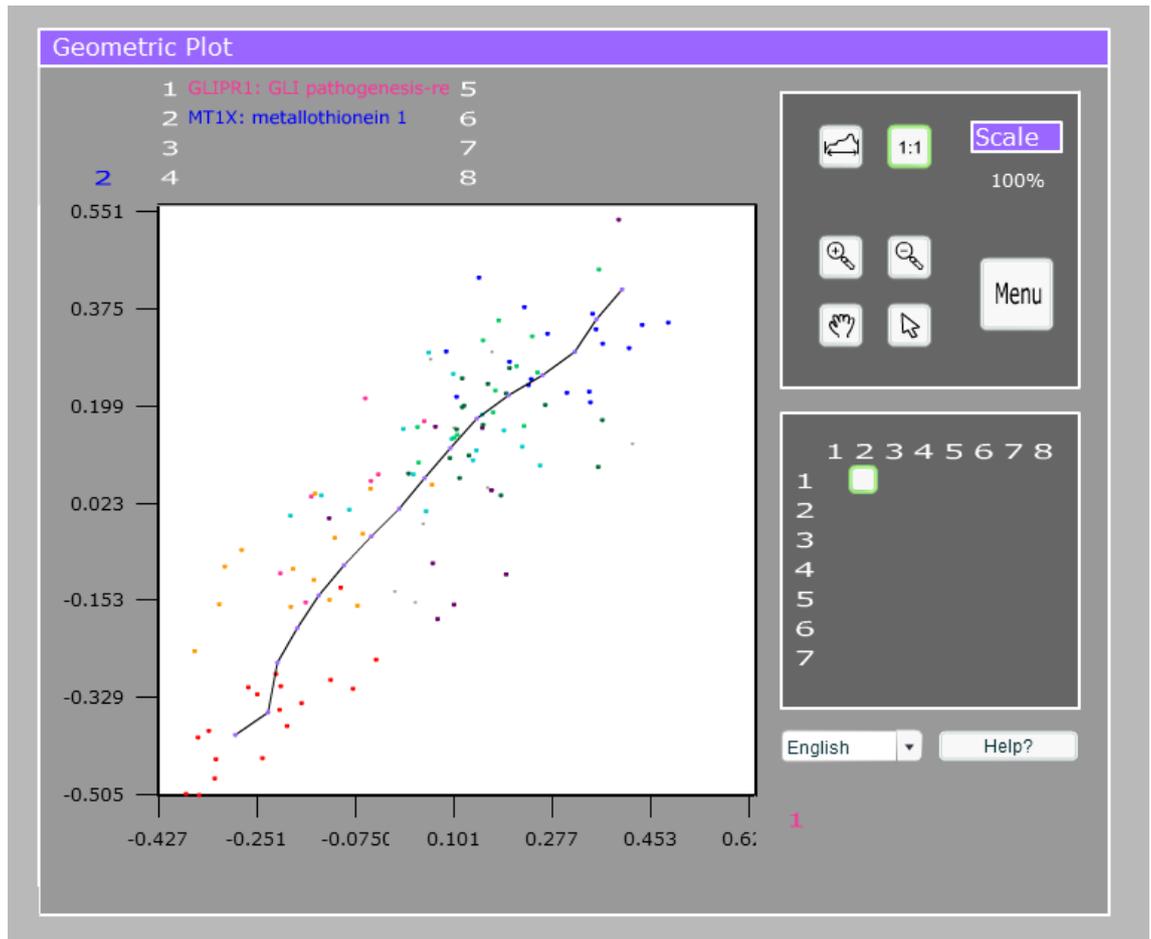


Figura 3.7 Vista del applet donde muestra la relación entre expresiones de genes que han sido elegidos por el usuario en las cajas de selección (checkboxes) al lado de cada gen.

En [figura 3.7] se muestra la relación entre los genes GLIPR1 y MT1X. Estos genes pueden haber sido seleccionados en cualquiera de las anteriores interfaces.

4. Informe Técnico

4.1 Estructura del servidor

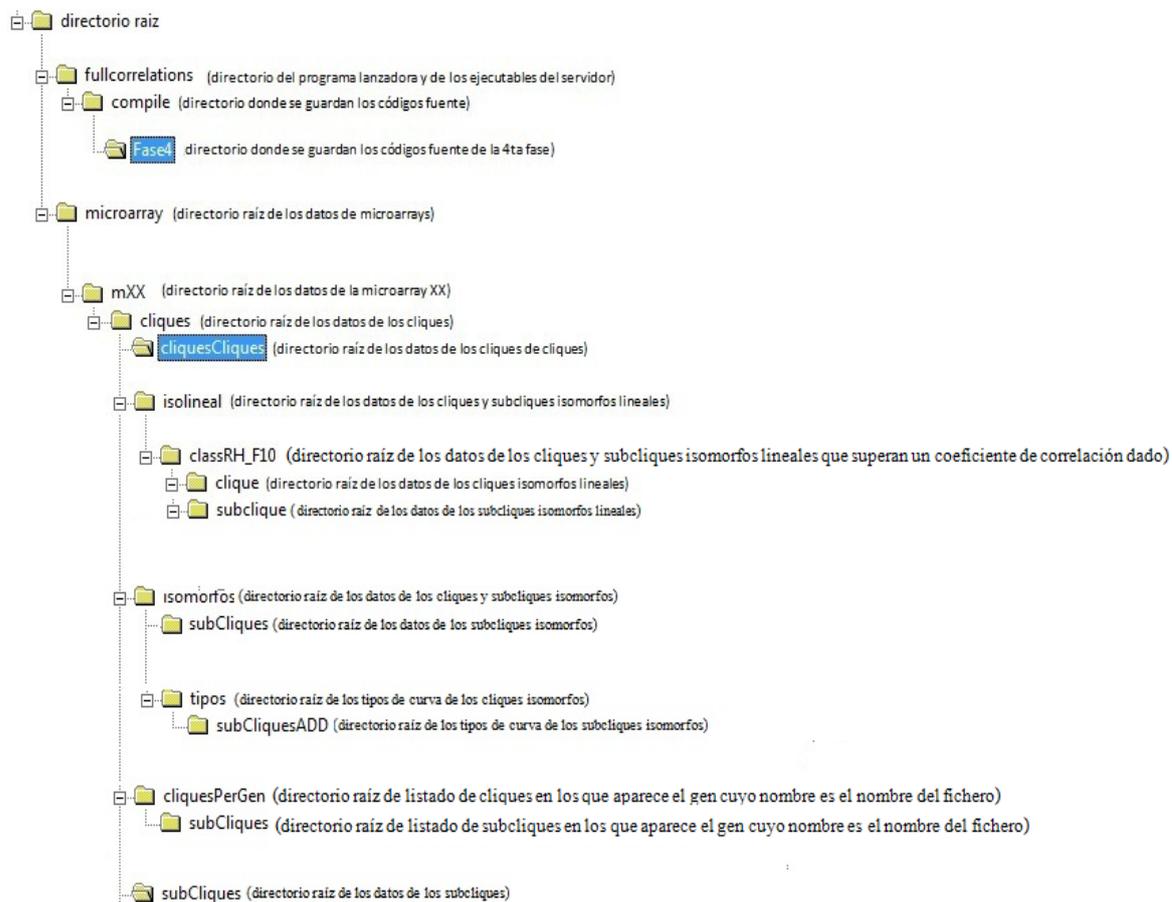
A continuación se procede a describir de qué forma está configurado el [servidor](#), la arquitectura de directorios y los ficheros de datos que contiene.

4.1.1 Estructura de directorios

La estructura de directorios definida en el [servidor](#), unifica los criterios de acceso a los datos por parte de todos los procesos que se puedan ejecutar. Por lo consiguiente, todos los procesos que intervienen en los diferentes cálculos saben donde se han de leer los datos y donde han de dejar los resultados para otros procesos.

La estructura de directorios en el servidor tiene una configuración fija, todos los directorios son relativos respecto al directorio raíz de la aplicación así pues no se utiliza nunca directorios absolutos para el acceso a datos y ejecución de procesos.

4.1.1.1 Representación de la arquitectura de directorios para el preproceso



Los directorios enmarcados en el cuadro azul (Fase4 y cliquesCliques) son los directorios creados por los cálculos que he añadido al preproceso y donde se guardan los datos generados por esos cálculos añadidos.

Dentro del directorio Fase4 se encuentran los códigos fuentes de los programas que se utilizan para generar los datos de esos nuevos cálculos del preproceso y que paso a describir.

4.1.1.1.1.- Ficheros de entrada utilizados como entrada para los nuevos cálculos del preproceso

En la carpeta cliques:

matrix.gr: Fichero que muestra las relaciones no lineales entre los genes de la microarray. En cada línea se muestran los genes que mantienen una relación no lineal con el gen cuyo identificador es el número de línea.

Connexions.ed : Fichero que muestra todos los cliques encontrados. El formato es X,X>Y. Donde X,X es una tupla de genes y “Y” es la relación que mantienen entre ellos. Cada línea representa un clique.

En la carpeta cliques->isolineal-> classRH_F10->clique:

procesosIso : Fichero que muestra los grupos de cliques isomorfos lineales clasificados por la tipología de sus aristas. Se presenta en una línea la tipología de las aristas de los cliques y a continuación, en las siguientes líneas, los grupos de cliques isomorfos lineales respecto a un clique dado pertenecientes a ese tipo de curvas. El primer clique del grupo será el clique con el que el resto de elementos del grupo son lineales.

procesosIso.g : Fichero que muestra los grupos de cliques isomorfos lineales clasificados por la tipología de sus aristas(**procesosIso**) representados por los genes que forman cada clique. Se presenta en una línea la tipología de las aristas de los cliques y a continuación, en las siguientes líneas, los grupos de cliques isomorfos lineales respecto a un clique dado pertenecientes a este tipo de curvas, representados por los genes que forman cada clique.

classList.lc: Fichero que muestra los diferentes tipos de curvas para todos los cliques. En la primera línea se muestra el número total de tipo de curvas y en las siguientes líneas los diferentes tipos de curvas.

En la carpeta cliques-> isomorfos->tipos:

allTipusCliques : Fichero que muestra todos los tipos de cliques isomorfos lineales y no lineales. Se presenta en una línea la tipología de las aristas de los cliques y a continuación, en la siguiente línea, los cliques isomorfos lineales y no lineales pertenecientes a este tipo de curvas.

4.1.1.1.2.- Programas desarrollados y los ficheros de salida generados en los nuevos cálculos del preproceso

todasFases: Programa que ejecuta todas las fases del preproceso. Serian 4 las fases que ejecutan. Los nuevos cálculos del preproceso se corresponderían con la fase 4.

Ficheros de entrada: En la fase 4 los ficheros de entrada son los ficheros ya disponibles generados en las anteriores fases.

Parámetros:

- 1.- identificador de la microarray.
- 2.- *coeficiente de correlación usado como corto.*

Fichero de salida: En la fase 4 se generarán los ficheros de los nuevos cálculos.

procesosIso_c: Fichero temporal. Muestra *procesosIso* sin la clasificación por la tipología de sus aristas. Mediante línea de comandos se eliminan todas las líneas que contengan la palabra clave “*curv*” del fichero *procesosIso*(*grep -v "curv" ../procesosIso > ../procesosIso_c*). En cada línea se muestran los grupos de cliques isomorfos lineales respecto a un clique dado

isolin2matrix: Convierte los datos de *procesosIso_c* a un fichero con el mismo formato que *matrix.gr*.

Ficheros de entrada: *procesosIso_c* , *connexions.ed*.

Parámetros:

1. Ruta del fichero de entrada *procesosIso_c*
2. Ruta de la carpeta del fichero de entrada *connexions.ed* y de los ficheros de salida *table_cliques.et* y *matrix_cliques*

Fichero de salida:

table_cliques.et: Fichero en el que se guardan los identificadores de los cliques de los grupos de cliques isomorfos lineales de *procesosIso_c* de manera que ahora los identificadores de los cliques de los grupos de cliques isomorfos lineales sean consecutivos. En cada línea se muestra el identificador del clique de los grupos de cliques isomorfos lineales que será reemplazado por el número de línea.

matrix_cliques.gr: Fichero en el que se encuentran los grupos de cliques de *procesosIso_c* con los nuevos valores numéricos asignados a los cliques.

MACE: Encuentra los cliques de los *grupos isomorficos lineales*, es decir los *cliques de cliques*.

Fichero de entrada: *matrix_cliques.gr*

Parámetros:

1. M: Un clique maximal es aquel clique que no esta incluido en ningún otro clique. Para nuestro uso a este parámetro se le indica, mediante la letra M, hallar solo cliques maximales.
2. -l 3: En este parámetro se indica, mediante -l, que los cliques han de tener un tamaño mínimo de número de nodos. Para nuestro uso este parámetro se le indicará un mínimo de 3 nodos.
3. Ruta del fichero de entrada *matrix_cliques.gr*
4. Ruta donde guarda el fichero de salida *out_cdc.gr*

Fichero de salida:

out_cdc.gr: Fichero temporal. Fichero con los *cliques de cliques*. El formato del fichero esta representado por una secuencia de valores numéricos que corresponden con los vértices del clique. Cada línea del fichero de salida es un clique de cliques. Los cliques que conforman el clique de cliques en cada línea están separados por “ “.

recoverCliquesValues : Recupera los identificadores originales de los cliques para el fichero de cliques de cliques.

Ficheros de entrada: *out_cdc.gr*, *table_cliques.et*

Parámetros:

- 1.-Ruta del fichero de entrada *table_cliques.et*
- 2.-Ruta del fichero de salida *cliques_cliques_c*

Fichero de salida:

cliques_cliques_c : Cliques de cliques con el formato del fichero procesosIso con los cliques con su identificador original. Los cliques que conforman el clique de cliques en cada línea están separados por “,”.

clasificar : Clasifica los cliques de cliques por la tipología de sus aristas.

Ficheros de entrada: *procesosIso*, *cliques_cliques_c*.

Parámetros:

1. Ruta del fichero de entrada *procesosIso*.
2. Ruta del fichero de entrada *cliques_cliques_c*.
3. Ruta del fichero de salida *cliques_cliques*.

Fichero de salida:

cliques_cliques : Los cliques de cliques clasificados por la tipología de sus aristas.

conversor : Encuentra los genes de los cliques que componen *cliques_cliques* para guardar los genes de cada clique por su nombre.

Fichero de entrada: *Connexions.ed*, *cliques_cliques*

Parámetros:

1. Ruta del fichero de entrada *Connexions.ed*
2. Ruta del fichero de entrada *cliques_cliques*
3. Ruta del fichero de salida *cliques_cliques_genes*

Fichero de salida:

cliques_cliques_genes : Los cliques de cliques representados por los genes que conforman los cliques.

converterT : Encuentra los genes que componen los cliques isomorfos lineales y no lineales. (*allTipusCliques*) para guardar los genes de cada clique por su nombre.

Ficheros de entrada: *Connexions.ed* ,*allTipusCliques*

Parámetros:

- 1.- Ruta del fichero de entrada *Connexions.ed*.
- 2.- Ruta del fichero de entrada *allTipusCliques*.
- 3.- Ruta del fichero de salida *allTipusCliques_genes*.

Fichero de salida:

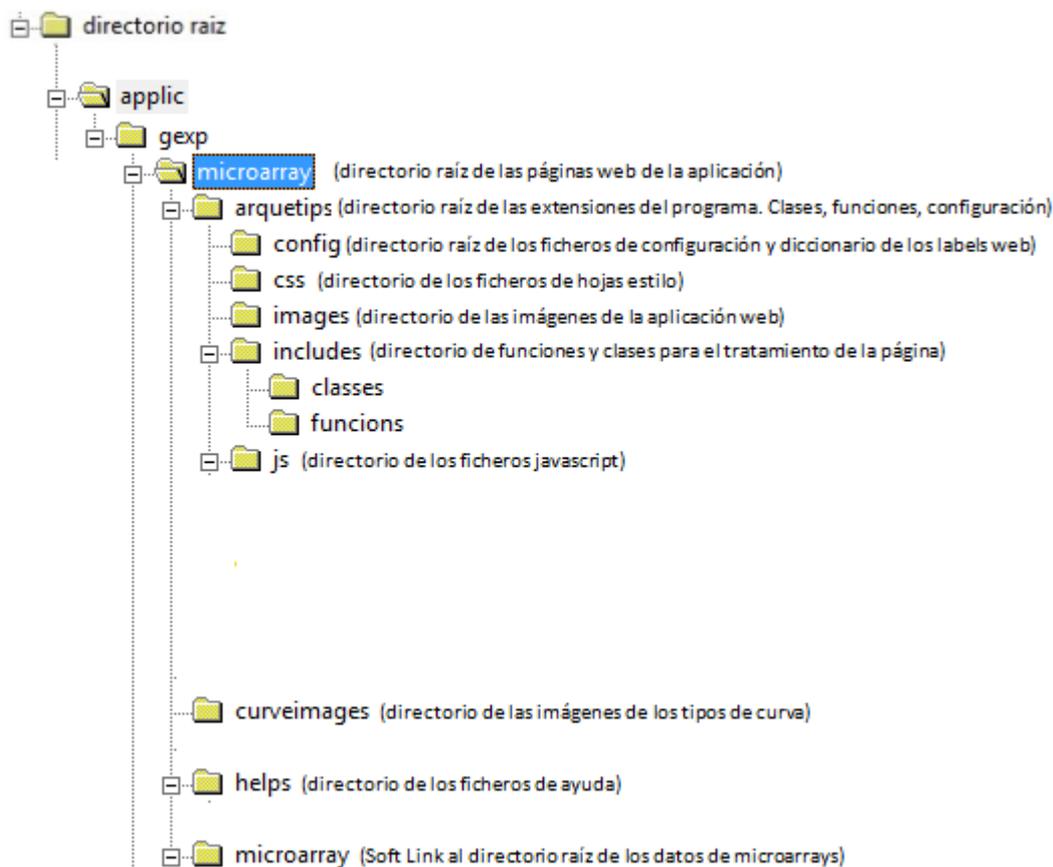
allTipusCliques_genes : Los tipos de cliques isomorfos lineales y no lineales representados por los genes que conforman los cliques.

Se eliminan los siguientes ficheros dado que ya han cumplido su función y no se volverán a usar:

procesosIso_c

out_cdc.gr

4.1.1.2 Representación de la arquitectura de directorios para la aplicación web



4.1.1.2.1.- PHPs desarrollados para la Aplicación web

cliquescliques.php: Pagina web que muestra los cliques de cliques clasificados por la tipología de sus aristas.

Fichero de entrada: *cliques_cliques* y *classList.lc*.

cliquescliquesdetail.php: Pagina web que muestra los cliques de cliques representados por los genes que conforman los cliques clasificados por la tipología de sus aristas

Fichero de entrada: *cliques_cliques_genes*

Parámetros:

- Recibe la tipología de las aristas de los cliques que se mostrarán.

cliquesIsolineales.php: Pagina web que muestra los grupos de cliques isomorfos lineales clasificados por la tipología de sus aristas

Fichero de entrada: *classList.lc* y *procesosIso*.

cliquesIsolinealedetails.php: Pagina web que muestra los grupos de cliques isomorfos lineales representados por los genes que conforman los cliques clasificados por la tipología de sus aristas

Fichero de entrada: *procesosIso.g*.

Parámetros:

- Recibe la tipología de las aristas de los cliques que se mostrarán.

cliques.php: Pagina web que muestra los cliques clasificados por la tipología de sus aristas.

Fichero de entrada: *allTipusCliques* y *classList.lc*

cliquesdetail.php: Pagina web que muestra los cliques representados por los genes que conforman los cliques clasificados por la tipología de sus aristas.

Fichero de entrada: *allTipusCliques_genes*

Parámetros:

- Recibe la tipología de las aristas de los cliques que se mostrarán.

5.-Conclusiones:

Después de haber realizado el proyecto, puede afirmarse que se han cumplido todos los objetivos que se habían propuesto inicialmente.

Se han conseguido pues obtener los ficheros que se deseaban en el preproceso y se ha conseguido diseñar una interfaz web donde se muestran los cliques de cliques, los grupos de cliques isomorficos lineales y los cliques con relaciones no lineales de alta curvatura.

1. En el preproceso:
 - a. Se han obtenido los cliques de cliques a partir de los grupos de cliques isomorficos lineales.

Interfaz Web para mostrar las relaciones de expresión entre grupos de genes

- i. Se dio el formato adecuado para que el MACE pudiera entender y recoger de los grupos de cliques isomorficos lineales y poder hallar los cliques de cliques.
 - b. Se clasificaron los cliques de cliques.
 - i. Se consiguieron los ficheros de los cliques de cliques clasificados teniendo en cuenta la tipología de sus aristas.
 - ii. Se dio el formato a los ficheros de cliques de cliques en los que se les representaba por los genes que los conformaban.
 - c. Se crearon los datos, ficheros y carpetas dotando a la interfaz web de los datos necesarios para que realice los menos cálculos posibles.
2. En el desarrollo de la interfaz web:
 - a. Se mostraron los cliques de cliques.
 - i. Las diferentes tipologías halladas para los cliques de cliques se muestran gráficamente.
 - ii. Se hizo un cuadro estadístico para cada tipología.
 - iii. Los cliques de cliques son mostrados para cada tipología.
 - iv. Se realiza la llamada a la interfaz del servidor que muestra en detalle la relación entre las expresiones de los genes elegidos por el usuario.
 - b. Se mostraron los grupos de cliques isomorficos lineales.
 - i. Las diferentes tipologías halladas para los grupos de cliques isomorficos lineales se muestran gráficamente.
 - ii. Se hizo un cuadro estadístico por cada tipo de tipología.
 - iii. Los cliques isomorficos lineales pertenecientes a cada tipología son mostrados.
 - iv. Se realiza la llamada a la interfaz del servidor que muestra en detalle la relación entre las expresiones de los genes elegidos por el usuario.
 - c. Se mostraron los cliques.
 - i. Se muestran solo cliques que tienen un mínimo del 35% de relaciones no lineales que no son de baja curvatura.
 - ii. Las diferentes tipologías halladas para los grupos de cliques isomorficos lineales se muestran gráficamente.
 - iii. Se hizo un cuadro estadístico por cada tipo de tipología.
 - iv. Se realiza la llamada a la interfaz del servidor que muestra en detalle la relación entre las expresiones de los genes elegidos por el usuario.

Este proyecto me ha resultado muy interesante pues me ha permitido afianzar mis conocimientos y habilidades que he ido adquiriendo durante mi formación universitaria. He podido adquirir conocimientos propios de la biomedicina, un campo que no se ve en la informática. La parte que me ha resultado más

interesante ha sido tanto la parte del preproceso, por afianzar mis conocimientos de programación en el lenguaje C, como la creación de la interfaz web, por obligarme a adquirir conocimientos en un lenguaje como php del que había visto más bien poco durante la carrera universitaria. Por otro lado, la parte que me ha resultado más tediosa ha sido el análisis de los datos iniciales, ya que había disponible mucha información y me resultó complicado entender todos los datos de los que tenía que partir.

6.- Referencias

1. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R: **NCBI GEO: mining millions of expression profiles – database and tools.** *Nucleic Acids Res* 2005, 33:D562-566.
2. Antonov AV, Tetko IV, Mewes HW: **A systematic approach to infer biological relevance and biases of gene network structures.** *Nucleic Acids Research* 2006, 34(1):e6.
3. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M, et al.: **ArrayExpress – a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Res* 2005, 33:D553-555.
4. Burgarella S, Cattaneo D, Pinciroli F, Masseroli M: **MicroGen: a MIAME compliant web system for microarray experiment information and workflow management.** *BMC Bioinformatics* 2005, 6(Suppl 4):S6.
5. Tarraga J, Medina I, Carbonell J, Huerta-Cepas J, Minguéz P, Alloza E, Al-Shahrour F, Vegas-Azcarate S, Goetz S, Escobar P, et al.: **GEPAS, a web-based tool for microarray data analysis and interpretation.** *Nucleic Acids Res* 2008, 36:W308-314.
6. J.M. Basart, “**Grafs: fonaments i algorismes.**” *Manuals de la UAB, 13. Publicacions de la UAB, ISBN 84-490-1420-4, 1994.*
7. <http://research.nii.ac.jp/~uno/code/mace.htm>
8. Huerta M, Cedano J, Querol E. : **Analysis of nonlinear relations between expression profiles by the principal curves of oriented-points approach.** *J Bioinform Comput Biol.* 2008 Apr;6(2):367-86.
9. Cedano J, Huerta M, Querol E.:**NCR-PCOPGene: An Exploratory Tool for Analysis of Sample-Classes Effect on Gene-Expression Relationships.***Adv Bioinformatics.* 2008:789026. *Epub* 2008 Dec 10.
10. Huerta M., Cerdano J., Peña D., Rodríguez A. y Querol E: **PCOPGene-Net: Holistic Characterisation of cellular states from microarray data based on continuous and non-continuous analysis of gene-expression relationships.** *BCM Bioinformatics,* 2009.
11. Cedano J, Huerta M, Estrada I, Balllllosera F, Conchillo O, Delicado P, Querol E.: **A web server for automatic analysis and extraction of relevant biological knowledge.***Comput Biol Med.* 2007 Nov;37(11):1672-5. *Epub* 2007 May 25.
12. Delicado, P. and Huerta, M.: **Principal Curves of Oriented Points: Theoretical and computational improvements.***Computational Statistics* 18, 293-315.(2003)

13. O'Madadhain J, Fisher D, Smyth P, White S, Boey YB: **Analysis and visualization of network data using JUNG**. *Journal of Statistical Software* 2005, VV:1-35.

Resum

El projecte que es presenta en aquesta memoria està compost per un preproces i una aplicació web. L'objectiu del treball realitzat és mostrar les relacions d'expressió entre grups de gens. A més de dissenyar l'aplicació web en : <http://revolutionresearch.uab.es> que permetrà l'estudi dels resultats obtinguts.

Resumen

El proyecto que se presenta en esta memoria esta compuesto por un preproceso y una aplicación web. El objetivo del trabajo realizado es mostrar las relaciones de expresión entre grupos de genes. Además de diseñar la aplicación web en: <http://revolutionresearch.uab.es> que permitirá el estudio de los resultados obtenidos

Summary

The Project introduced in this report is composed by a preprocessor and a web application. The aim of the work performed is to show relationships between groups of genes expression. In addition to designing the Web application: <http://revolutionresearch.uab.es> that allow the study of the results obtained.