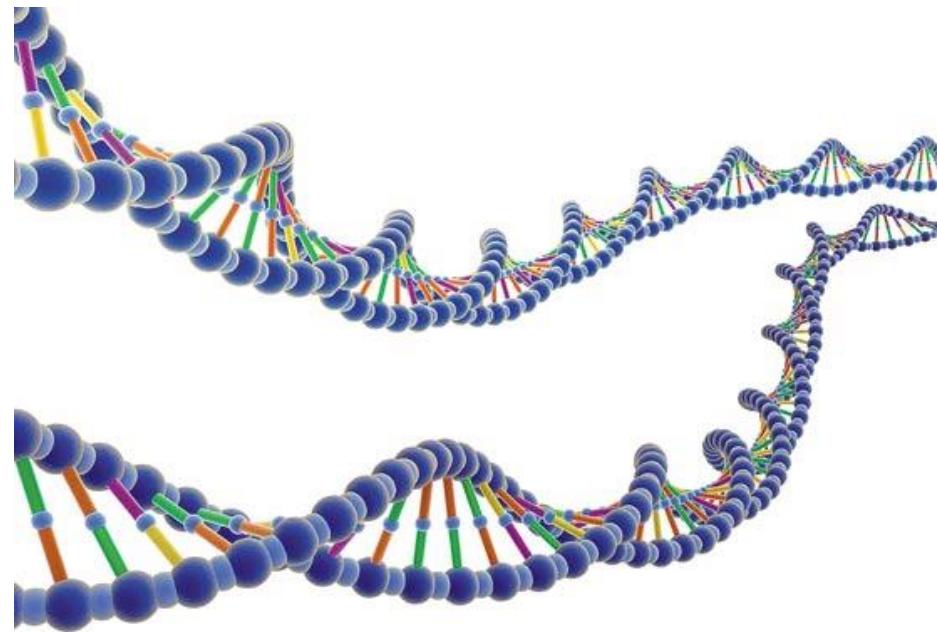


Interfaz web para mostrar los genes según las relaciones que mantienen sus expresiones



PROYECTO REALIZADO POR:

Luis Alberto Hernandez Gracia

BAJO LA DIRECCION DE:

Jordi Gonzalez, Mario Huerta

1. Introducción

2. Objetivos

3. Fases

4. Resultados

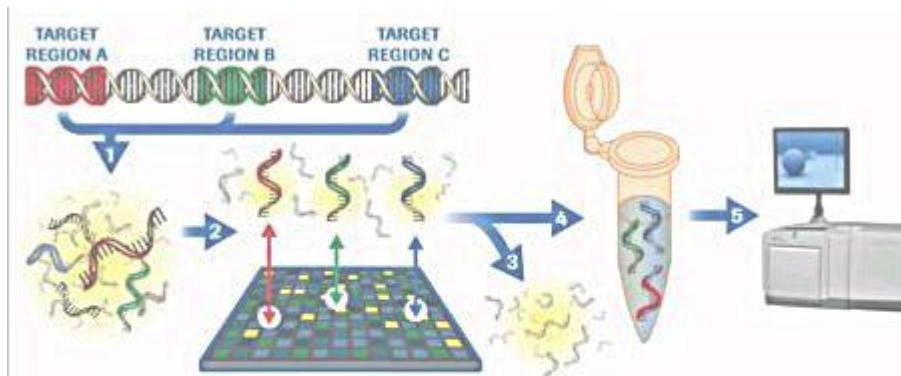
5. Conclusiones

Expresión de los genes.

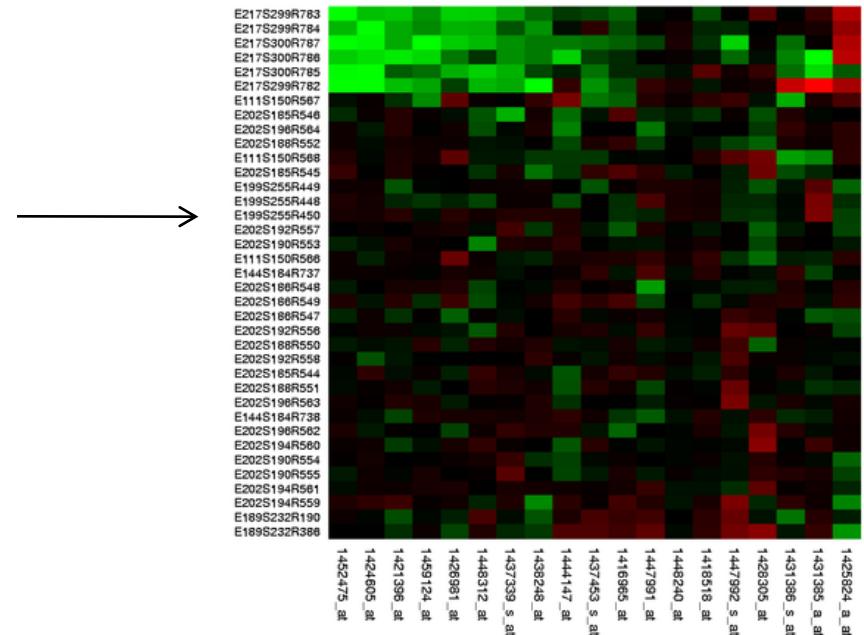
- Cuando los genes se expresan se sintetiza la proteína correspondiente a cada gen a partir de la información codificada en los ácidos nucleicos del ADN.
- Las proteínas sintetizadas dan lugar a las funciones celulares. En algunos casos específicas de cada tejido permitiendo el correcto funcionamiento del organismo.
- La interacción entre las proteínas sintetizadas por genes que se han expresado a la vez puede determinar tanto procesos biológicos funcionales como procesos patológicos.

Estudio de niveles de expresión entre genes mediante la técnica de microarrays

En la microarray, fragmentos de ADN se insertan en pequeñas casillas en forma de matriz sobre una superficie sólida. Los fragmentos de ADN se hibridan y mediante escaneo óptico se obtiene el nivel de expresión de los genes.

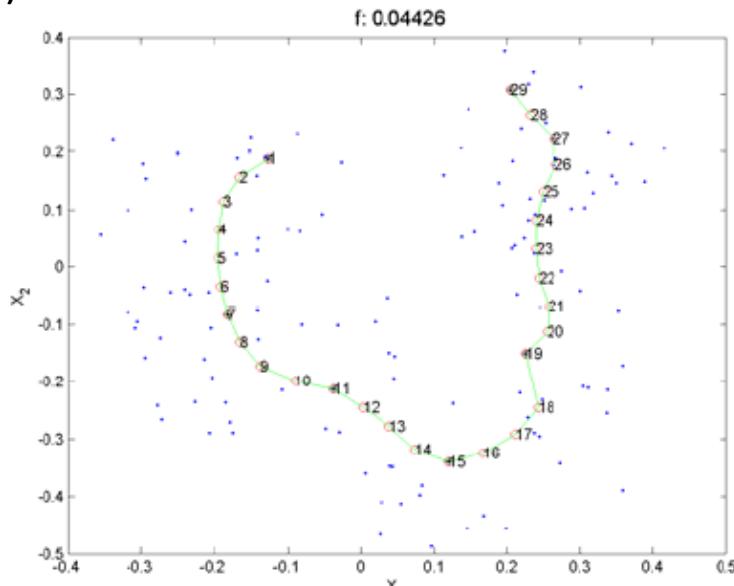


1. The genomic DNA sample is fragmented.
 2. The sample is hybridized to a custom NimbleGen Sequence Capture array.
 3. Unbound fragments are removed.
 4. The target-enriched pool is eluted and amplified.
 5. The enriched sample is ready for high-throughput sequencing, such as with a 454 GS FLX instrument.



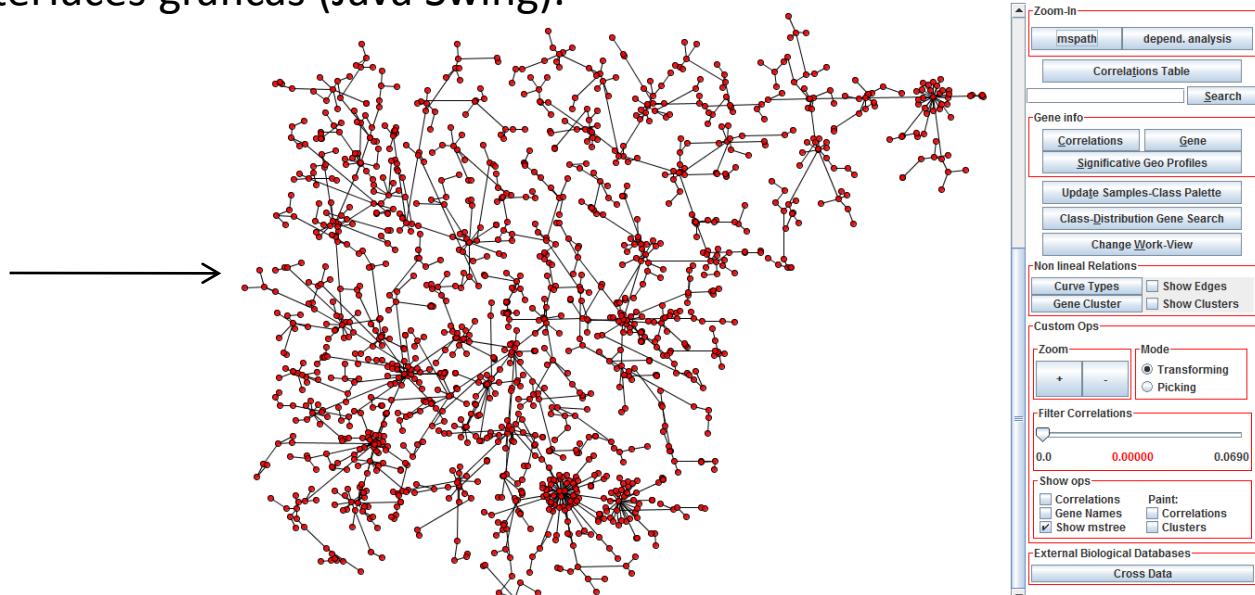
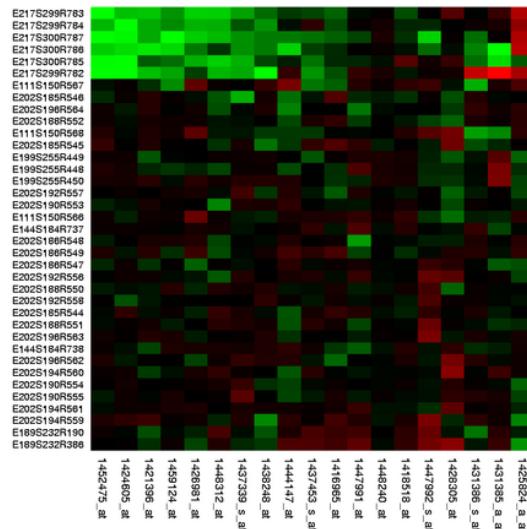
Relaciones de expresión entre genes.

- La relación de expresión entre dos genes describe el nivel de expresión de un gen respecto al nivel de expresión del otro gen.
- La relación de expresión puede ser lineal o no lineal.
- Estudio de las relaciones de expresión no lineales mediante las Principal Curves of Oriented Points (PCOP).



PCOPGene-Net

- PCOPGene-Net es una aplicación web implementada en Java. Permite estudiar las relaciones de expresión de los genes a partir de los datos de una microarray.
 - Los datos de las microarrays que usa PCOPGene-Net son tratados previamente por un preproceso en el servidor revolutionresearch.uab.es del IBB.
 - El análisis de los datos de la microarray se realiza de manera gráfica mediante un grafo interactivo (JUNG) y unas interfaces gráficas (Java Swing).



1. Introducción

2. Objetivos

3. Fases

4. Resultados

5. Conclusiones

¿Qué se busca en este proyecto?

- Facilitar la investigación de las relaciones de expresión entre genes a los usuarios del applet PCOPGene-Net.

¿Qué importancia tienen las relaciones de expresión entre genes?

- Aportan mucha información sobre el desarrollo de los procesos biológicos que permiten el funcionamiento de los organismos así como de los procesos patológicos, como por ejemplo enfermedades con gran componente genético como algunos tipos de cáncer.

¿Qué ventaja tiene el uso de la técnica de las microarrays para el estudio de las relaciones de expresión entre genes?

- La técnica de las microarrays permite estudiar la relación de expresión entre miles de genes a la vez.

Objetivos generales:

Facilitar la tarea investigadora a los usuarios de PCOPGene-Net:

- Identificar sobre PCOPGene-Net los genes de una microarray según la relación de expresión que mantienen con el resto de genes.
- Mejorar el proceso de identificación en PCOPGene-Net de los diferentes tipos de relaciones de expresión entre los genes de una microarray.

Para cumplir los objetivos propuestos he definido las siguientes fases:

1. Generar una gráfica con la relación entre los clústeres de genes por los tipos de relaciones de expresión de cada gen con los diferentes tipos de relación de expresión, para mostrarla en PCOPGene-Net.
2. Identificación sobre el grafo interactivo de los genes que pertenecen cada clúster en PCOPGene-Net.
3. Mejorar la identificación de los diferentes tipos de relaciones de expresión no lineales en PCOPGene-Net.
4. Mejorar el rendimiento y uso de recursos de PCOPGene-Net.

1. Introducción

2. Objetivos

3. Fases

4. Resultados

5. Conclusiones

1. Generar una gráfica con la relación entre los clústeres de genes por los tipos de relaciones de expresión de cada gen con los diferentes tipos de relación de expresión, para mostrarla en PCOPGene-Net.

1.1 Modificación del preprocesso de clustering de genes por los tipos de relaciones de expresión de cada gen.

Modificar el preprocesso de clustering basado en llamadas a funciones de la librería de clustering de CLUTO por llamadas al programa *vcluster* de CLUTO para poder obtener la gráfica.

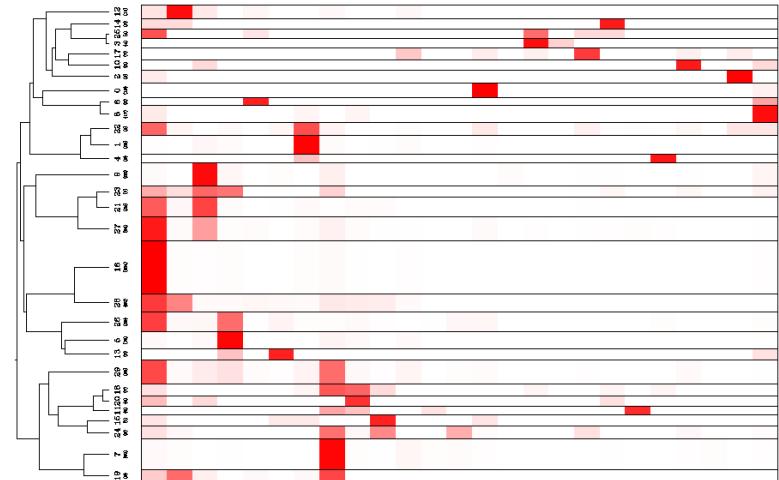
1. Identificar los parámetros de entrada usados en las llamadas a funciones.
2. Construir las llamadas a *vcluster* con los parámetros encontrados.
3. Comprobar que el clustering realizado con *vcluster* es igual al realizado por el preprocesso de clustering anterior.
4. Añadir las llamadas a *vcluster* al preprocesso que trata los datos de la microarrays que se utilizaran en PCOPGene-Net.

1. Generar una gráfica con la relación entre los clústeres de genes por los tipos de relaciones de expresión de cada gen con los diferentes tipos de relación de expresión, para mostrarla en PCOPGene-Net.

1.2 Tratamiento de las gráficas obtenidas.

- 1- Obtener dimensiones de la gráfica y tamaño de microarray
- 2- Recorte y rotación del árbol de jerarquía de los clústeres
- 3- Recorte y rotación de leyenda según tamaño de microarray
- 4- Elección de la leyenda a utilizar
- 5- Composición de la matriz gráfica con la leyenda

ACIRCULAR
A51-———ORTO-ORTO
A51-L2_-CERR-CERR
A51-L1p-ORTO-CERR
A51-L1p-CERR-ORTO
A51-L1p-CERR-CERR
A51-L1n-ORTO-CERR
A51-L1n-CERR-ORTO
A51-L1n-CERR-CERR
A51-L1n-ABIE-CERR
A48-———ORTO-ORTO-AALT
A48-L2_-CORT-CERR-AALT
A48-L2_-CERR-ORTO-AALT
A48-L2_-CERR-CERR-AALT
A48-L1_-ORTO-CERR-AALT
A48-L1_-CERR-CERR-AALT
A48-L1_-ORTO-ORTO-AALT
A26-———ORTO-CORT-AALT
A26-L2_-CORT-CERR-AALT
A26-L2_-CERR-ORTO-AALT
A26-L1_-ORTO-CERR-AALT
A26-L1_-CERR-ORTO-AALT
A26-L1_-CERR-CERR-AALT



2. Identificación sobre el grafo interactivo de los genes que pertenecen a cada clúster en PCOPGene-Net.

2.1 Creación de la interfaz gráfica de identificación de los genes que pertenecen a cada clúster.

Nuevo menú que permita escoger un color para cada clúster para así identificar sobre el grafo interactivo a qué clúster pertenece cada gen. Para proporcionar dicha funcionalidad he planteado utilizar los siguientes elementos:

- Tres tablas dinámicas interactivas, una para cada tipo de clustering aplicado en el preprocesso.
- Desplegable que presenta la paleta de colores.
- Tres parejas de imágenes que representan el resultado de los diferentes tipos de clustering realizados en el preprocesso.
- Casillas de selección de los tres tipos de clustering.
- Botón de cierre del menú de coloreado de genes de cada clúster y botón de reseteo de las selecciones de colores.

2. Identificación sobre el grafo interactivo de los genes que pertenecen a cada clúster en PCOPGene-Net.

2.2 Modificación del proceso de coloreado de vértices del grafo interactivo.

- JUNG permite personalizar proceso de pintado de vértices.
- En PCOPGene-Net la clase personalizada es MyVertexPaintFunction.
- Se llama al método getFillPaint de MyVertexPaintFunction para cada gen del grafo.

SI coloreado de genes según el clúster al que pertenecen activado ENTONCES:

 Consultar tipo de clustering seleccionado.

 Consultar el archivo de resultado de clustering para saber que clúster le corresponde al gen.

 Retornar color asignado por usuario a ese clúster.

SINO SI coloreado de genes según el clúster al que pertenecen desactivado ENTONCES:

 SI el vértice está seleccionado ENTONCES

 Retornar color amarillo.

 SINO

 Retornar color rojo.

 FIN SI.

FIN SI.

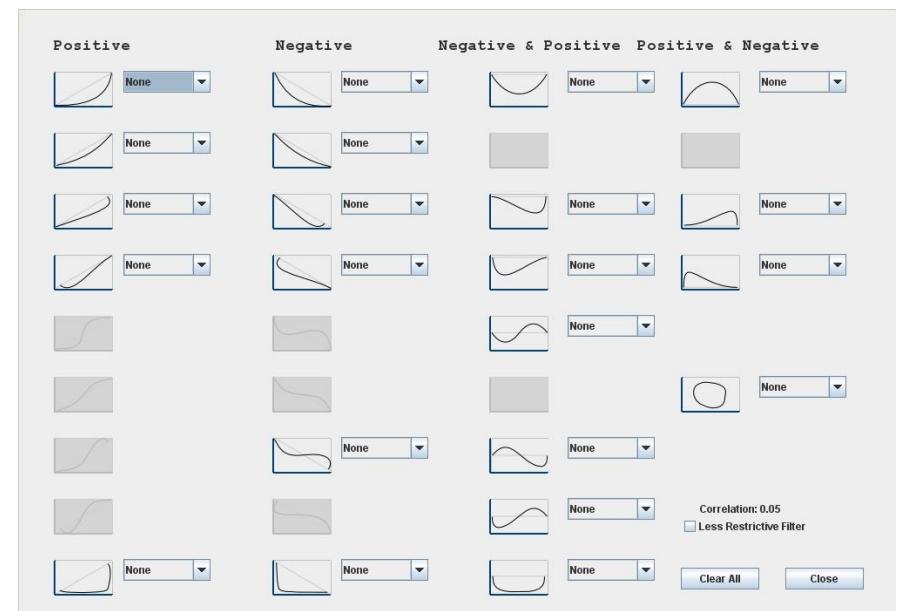
3. Mejorar la identificación de los diferentes tipos de relaciones de expresión no lineales en PCOPGene-Net.

3.1 Corrección de la interfaz gráfica de coloreado de las relaciones de expresión no lineales.

PCOPGene-Net dispone de un menú para asignar colores predefinidos a los 29 diferentes tipos de relaciones no lineales. Dicho menú carece de 2 funcionalidades que mejorarían la experiencia del usuario de PCOPGene-Net.

Nuevas funcionalidades a incorporar:

- Recordar la selección de colores.
- Reseteo de la selección de colores.



3. Mejorar la identificación de los diferentes tipos de relaciones de expresión no lineales en PCOPGene-Net.

3.2 Corrección y personalización del proceso de coloreado de aristas del grafo interactivo.

- JUNG permite personalizar proceso de pintado de aristas.
- En PCOPGene-Net la clase personalizada es MyEdgePaintFunction.
- MyEdgePaintFunction contiene errores:
 - Entre dos genes del grafo interactivo se pintan dos relaciones de expresión a la vez.
 - En el inicio de PCOPGene-Net se pintan relaciones de expresión que no forman parte del árbol de expansión mínima.

Estos errores se solucionan usando una estructura de control de las relaciones de expresión representadas por las aristas del grafo.

4. Mejorar el rendimiento y uso de recursos de PCOPGene-Net.

4.1 Eliminación del uso de pares de cadenas de tipo String que actúan como listas relacionadas por la posición de sus elementos.

4.2 Eliminación de bucles que usan StringTokenizer.

4.3 Declaración de variables estáticas en los procesos de coloreado.

1. Introducción

2. Objetivos

3. Fases

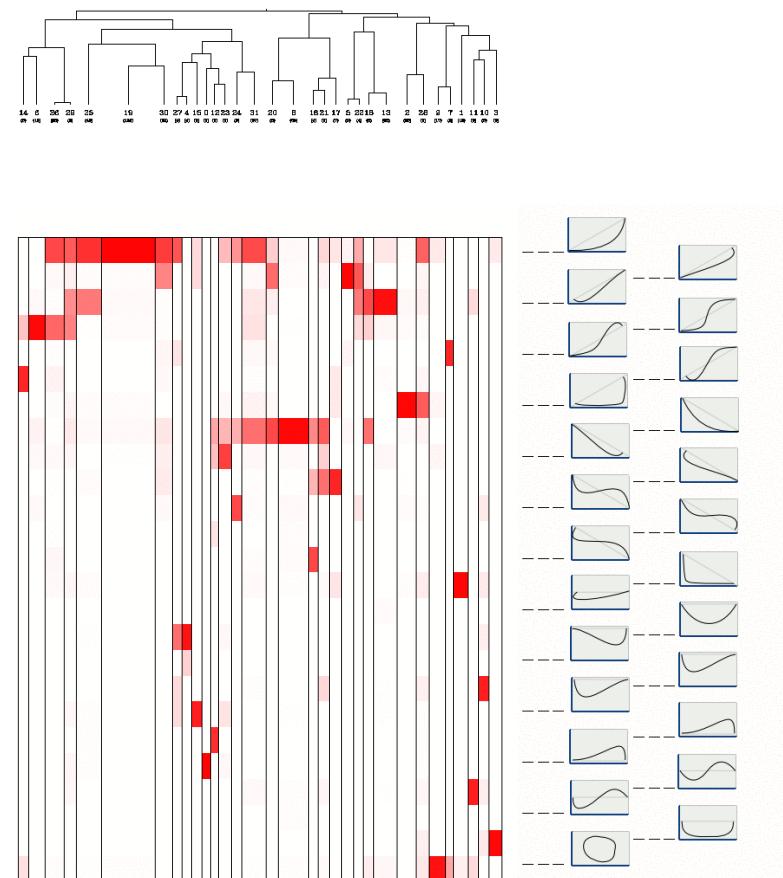
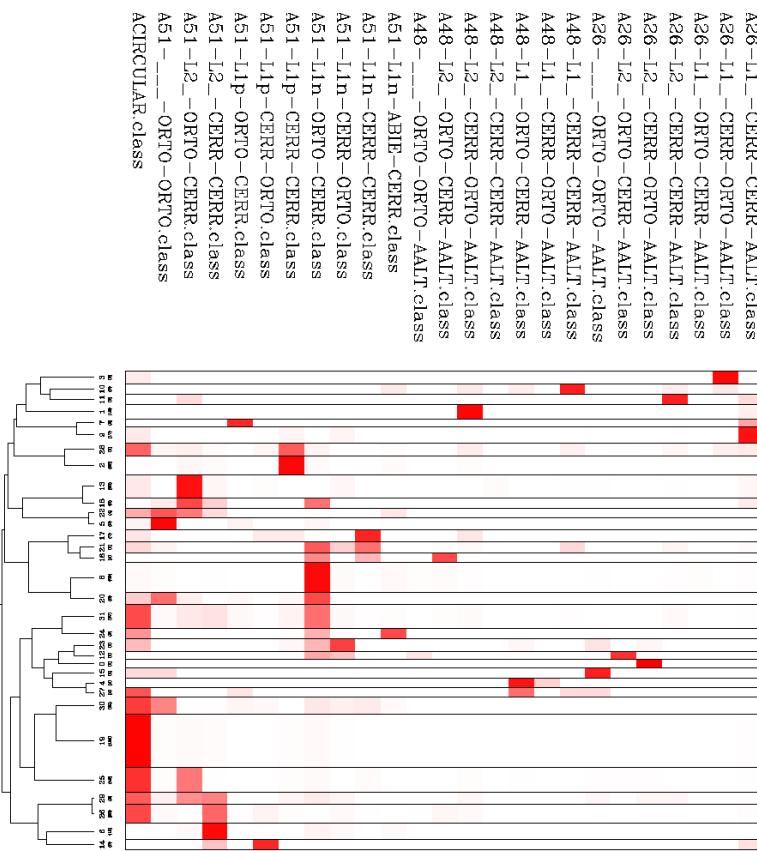
4. Resultados

5. Conclusiones

Resultados

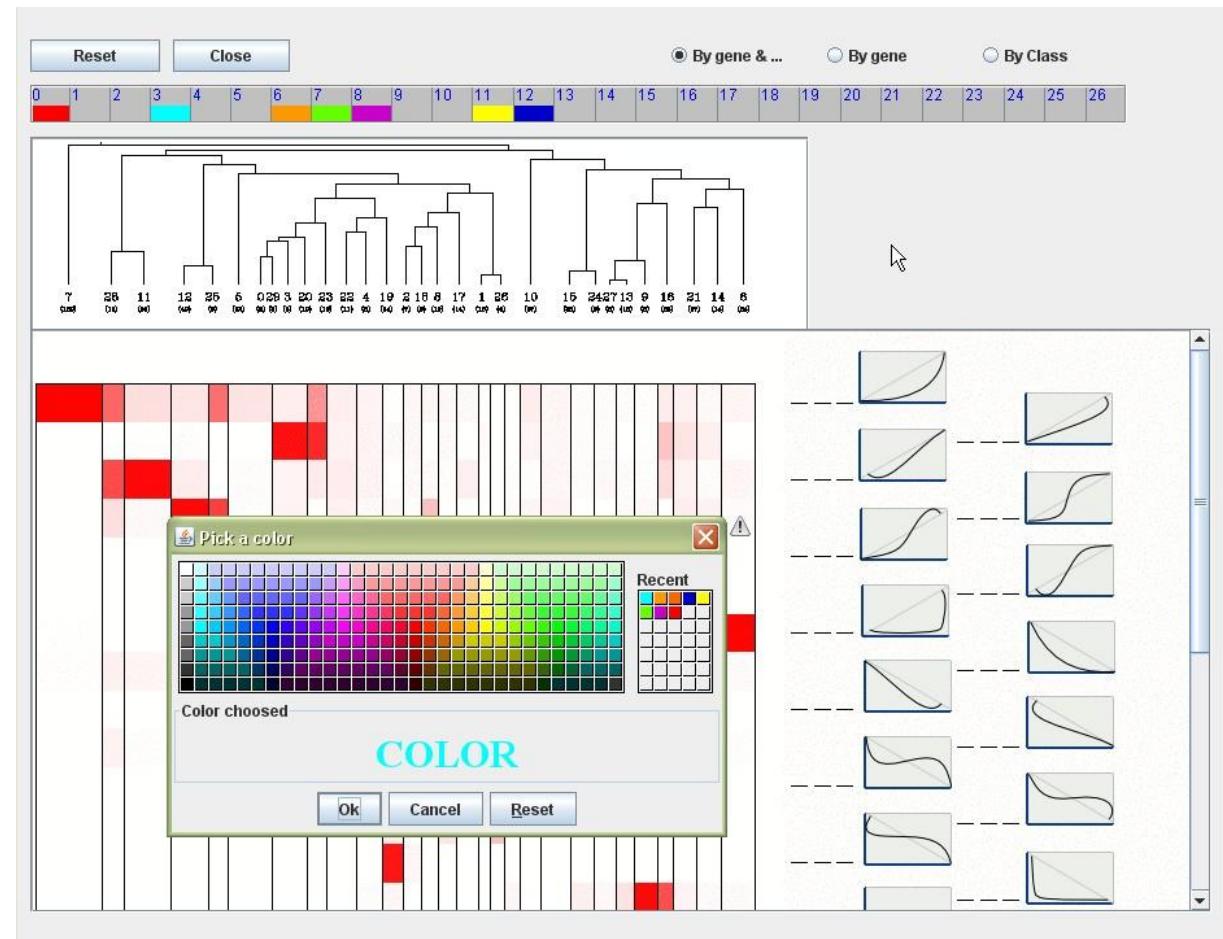
20/29

Resultados de generar una gráfica con la relación entre los clústeres de genes por los tipos de relaciones de expresión de cada gen con los diferentes tipos de relación de expresión



Resultados de identificación sobre el grafo interactivo de los genes que pertenecen a cada clúster en PCOPGene-Net

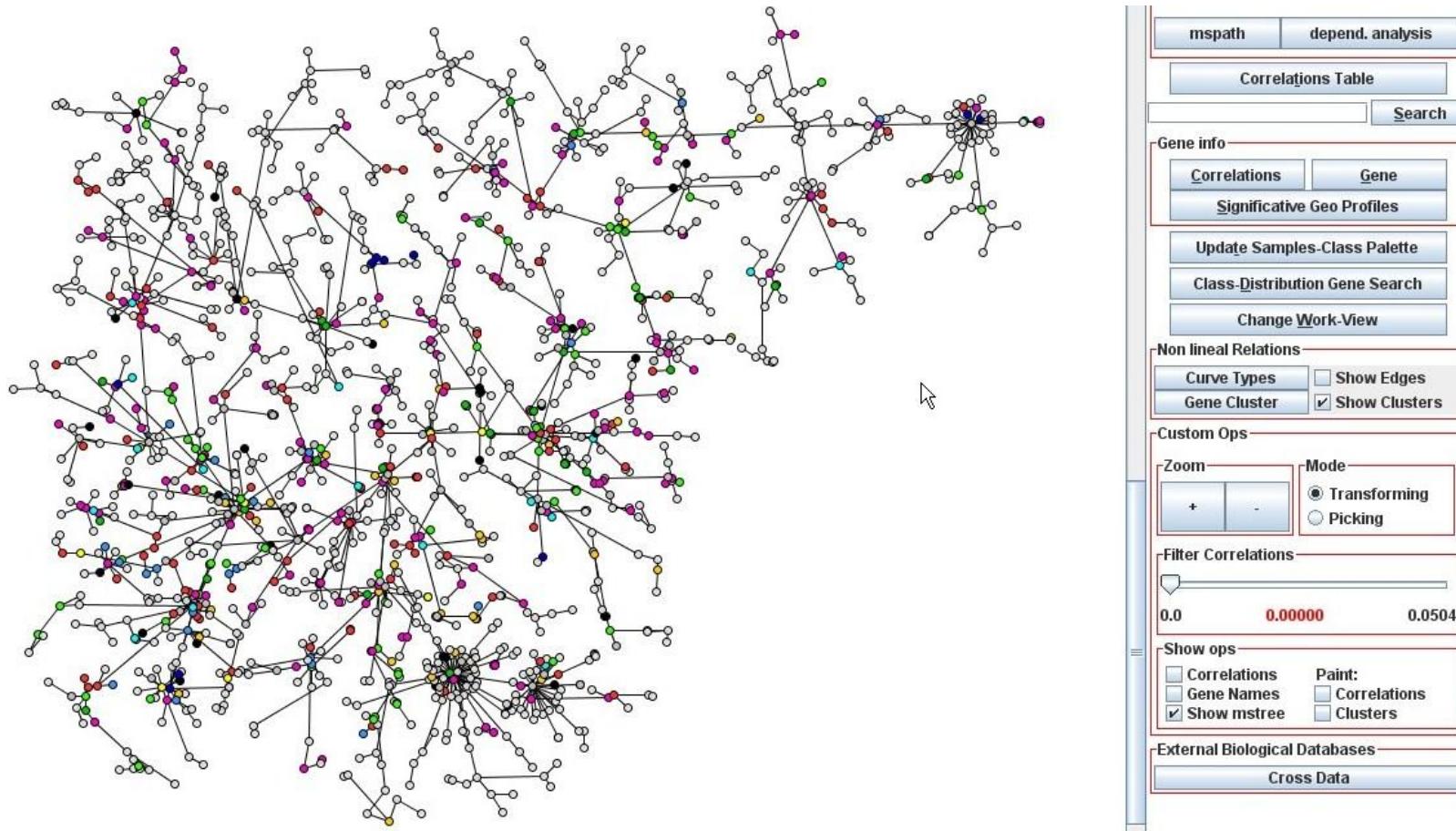
- Tabla dinámica interactiva
- Selector de color
- Gráficas de proceso de clustering
- Selectores de tipo de clustering
- Botón de reseteo de selección
- Botón de cierre del menú



Resultados

22/29

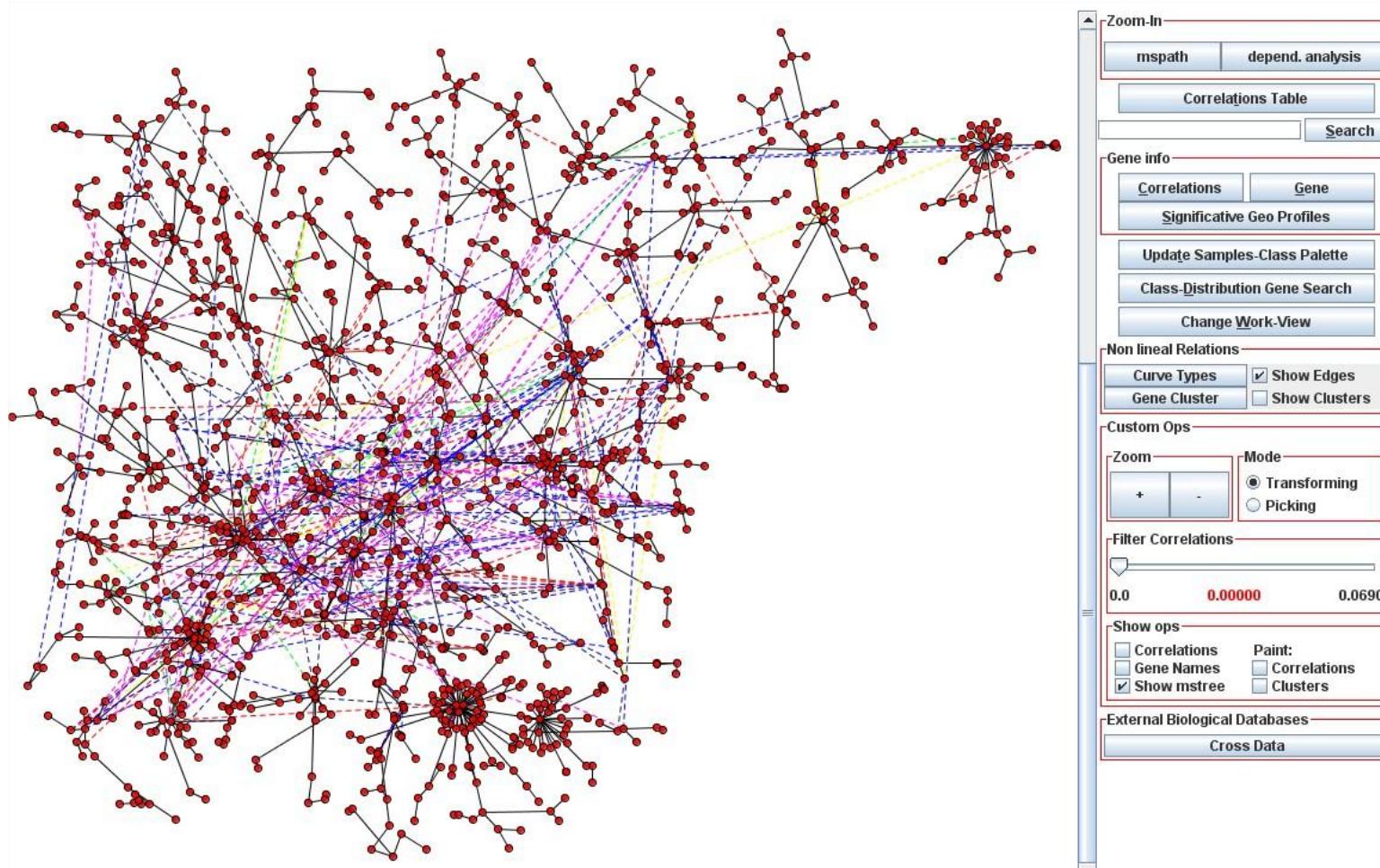
Resultados de identificación sobre el grafo interactivo de los genes que pertenecen a cada clúster en PCOPGene-Net



Resultados

23/29

Resultados de mejorar la identificación de los diferentes tipos de relaciones de expresión no lineales en PCOPGene-Net.

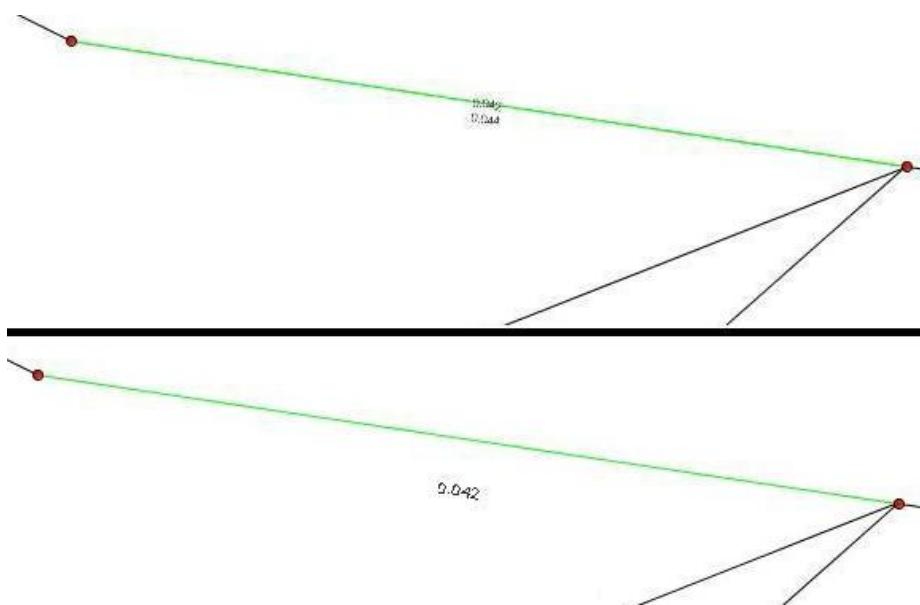


Resultados

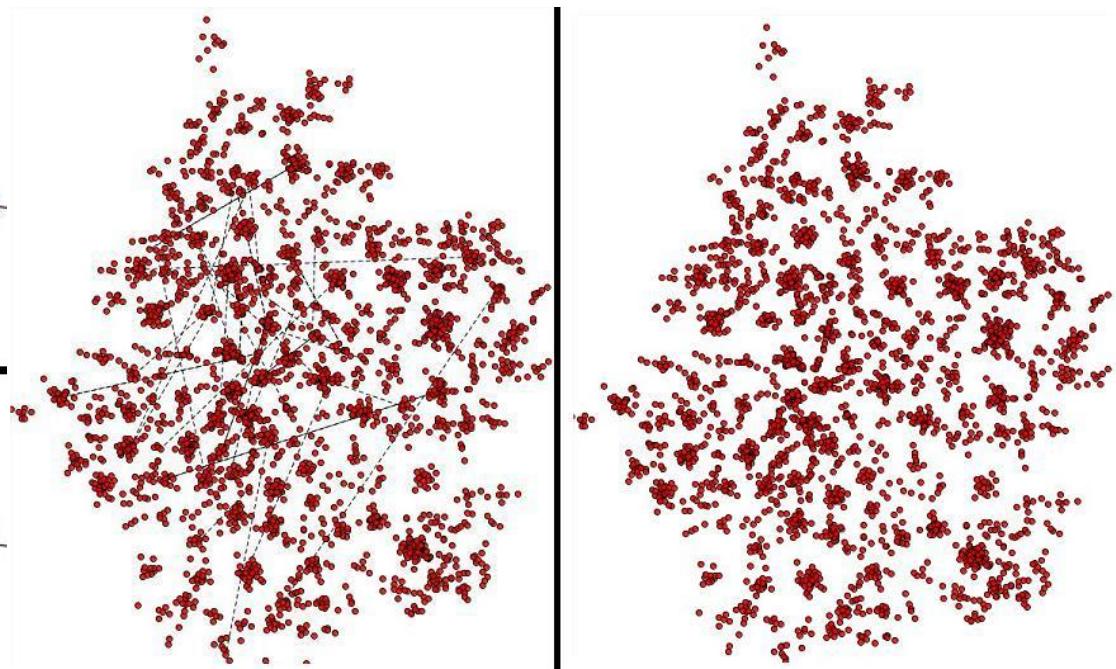
24/29

Resultados de mejorar la identificación de los diferentes tipos de relaciones de expresión no lineales en PCOPGene-Net.

Duplicado de aristas



Pintado incorrecto de aristas



Resultados de mejorar el rendimiento y uso de recursos de PCOPGene-Net.

Resultados con la microarray m17 de 1416 genes.

Memoria RAM	Versión anterior	Nueva versión
Inicio	32 MB	72 MB
Pintado de relaciones no lineales y sus grados	67 a 150+ MB	76 MB
Pintado de relaciones no lineales, grados y clústeres	--	82 MB
Uso de CPU	Versión anterior	Nueva versión
Inicio	48%	48%
Pintado de relaciones no lineales y sus grados	50 %	50%
Pintado de relaciones no lineales, grados y clústeres	--	51%
Tiempo de carga	Versión anterior	Nueva versión
Tiempo de carga del applet PCOPGene-Net	25093 ms	24678 ms

1. Introducción

2. Objetivos

3. Fases

4. Resultados

5. Conclusiones

- Gracias a la metodología aplicada se han cumplido todos los objetivos
- He comprobado que PCOPGene-Net funciona correctamente con microarrays de 664, 1416, 3000 y 5000 genes.
- Trabajo futuro:
 - Aplicar más optimizaciones que implican cambios profundos en la codificación de PCOPGene-Net.
 - Añadir nuevas herramientas a PCOPGene-Net que proporcionen aún más información sobre las relaciones de expresión entre genes.

- Cumplidos los objetivos propuestos a pesar de escasa documentación previa.
- La mejora y optimización de código ha resultado la tarea más gratificante.
- Descubrimiento del campo de la bioinformática y la genómica.
- Participación en un proyecto de investigación puntero en el IBB.

Delicado, P.(2001) Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, 77, 84-116.

Delicado, P. and Huerta, M. (2003): 'Principal Curves of Oriented Points: Theoretical and computational improvements'. *Computational Statistics* 18, 293-315.

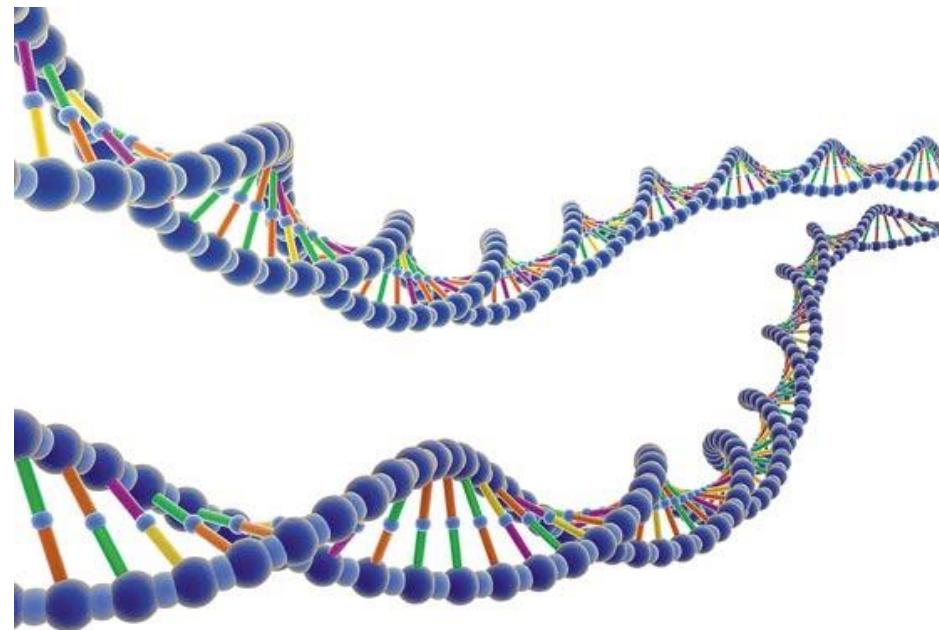
Cedano J, Huerta M, Estrada I, Ballllosera F, Conchillo O, Delicado P, Querol E. (2007) A web server for automatic analysis and extraction of relevant biological knowledge. *Comput Biol Med*. 37:1672-1675.

Huerta M, Cedano J, Querol E. (2008) Analysis of nonlinear relations between expression profiles by the principal curves of oriented-points approach. *J Bioinform Comput Biol*. 6:367-386.

Cedano J, Huerta M, Querol E. (2008) NCR-PCOPGene: An Exploratory Tool for Analysis of Sample-Classes Effect on Gene-Expression Relationships *Advances in Bioinformatics*, vol. 2008.

Huerta M, Cedano J, Peña D, Rodriguez A, Querol E. (2009) PCOPGene-Net: holistic characterisation of cellular states from microarray data base on continuous and non-continuos analysis og gene-expression relationships. *BMC Bioinformatics* 2009 May 9;10:138.

Interfaz web para mostrar los genes según las relaciones que mantienen sus expresiones

**PROYECTO REALIZADO POR:**

Luis Alberto Hernandez Gracia

BAJO LA DIRECCION DE:

Jordi Gonzalez, Mario Huerta