# Herramienta de Proyección y Agrupación de Micro Colección de Genomas

# **Herramienta MAGIC**

# Guía del Usuario



Herramienta MAGIC v2.1

15 de mayo de 2008

El Propósito de la Herramienta MAGIC	3
Información de Publicación	
Instalación de la Herramienta MAGIC	4
Vocabulario	
Instrucciones Básicas	6
Comenzar un Proyecto	7
Agregar Archivos al Proyecto	7
Cargar Archivos Tiff (Control R y Control G)	7
Cargar Lista de Genes (Control X)	8
Localizar Puntos	8
Distinga una señal de fondo y genere un archivo de expresión	<b>16</b>
Repita los pasos 1-6 para todas las condiciones experimentales	20
Manipule Datos	20
Agregar información de genes al archivo de expresión	22
Explorar datos	22
Filtrar datos	23
Calcular disimilitudes	24
Agrupe Genes	25
Automatización de Tareas	25
Comentarios Finales	<b>26</b>
Lista Completa de Opciones de la Herramienta MAGIC	<b>26</b>
Menú de Proyectos	<b>26</b>
Menú para Construir Archivo de Expresión	28
Menú de Expresión	<b>39</b>
Menú de Agrupación	48
Menú de Tareas	52
Créditos	53

# El Propósito de la Herramienta MAGIC

El propósito de la Herramienta MAGIC es permitir al usuario efectuar un proceso que empieza con una micro-colección de ADN de archivos de extensión tiff y que termina con información biológica significativa. Datos de hibridación comparativa (llamados chips de vidrio, en ingles denominados "glass chips") y datos de tipo Affymetrix son compatibles con la Herramienta MAGIC. Usted puede empezar con archivos de extensión tiff o archivos de expresión (hojas de cálculo con proporciones o niveles de expresión absolutos).

La herramienta MAGIC ha sido creada tomando en cuenta el nivel de habilidad del principiante pero no es un programa "tonto". De hecho, la Herramienta MAGIC ha sido diseñada para aclarar los algoritmos que se emplean en los procesos en lugar de ser una caja negra que produce resultados con poca información de entrada de la parte del usuario. La Herramienta MAGIC permite al usuario cambiar los parámetros de agrupación, cuantificación de datos, etc. Esta Guía del Usuario le demostrará cómo utilizar el software pero deja las explicaciones teóricas y el razonamiento matemático a la Guía del Instructor.

Se recomienda que los usuarios visiten los siguientes sitios relacionados (cuyo contenido está solamente disponible en inglés):

Página web de MAGIC: <a href="http://www.bio.davidson.edu/MAGIC">http://www.bio.davidson.edu/MAGIC</a>

Laboratorio en línea de la Herramienta MAGIC: <a href="http://gcat.davidson.edu/GCAT/workshop2/derisi\_lab.html">http://gcat.davidson.edu/GCAT/workshop2/derisi\_lab.html</a>

Tutoría de Agrupación: <a href="http://gcat.davidson.edu/DGPB/clust/home.htm">http://gcat.davidson.edu/DGPB/clust/home.htm</a>

GCAT: http://www.bio.davidson.edu/GCAT

Curso de Genómica: http://www.bio.davidson.edu/genomics

Soporte para la Herramienta MAGIC es proporcionado por los autores y estudiantes asistentes (con el apoyo de la NSF). Usted puede enviar un correo a <a href="mailto:magictool.help@gmail.com">magictool.help@gmail.com</a> o <a href="mailto:laheyer@davidson.edu">laheyer@davidson.edu</a> para obtener dicho soporte. Usted también puede enviar un correo a la "listserv" de GCAT para obtener ayuda, ya que hay varios usuarios de la Herramienta MAGIC en esta lista. Visite <a href="http://www.bio.davidson.edu/projects/gcat/GCAT-L.html">http://www.bio.davidson.edu/projects/gcat/GCAT-L.html</a> para obtener más información acerca de esta lista "listserv" de GCAT.

Se ha incorporado las siguientes funciones en la versión 2.1 de la Herramienta MAGIC:

- Los usuarios pueden mover cuadrículas múltiples a la vez con la tecla "Shift"
- Los usuarios pueden combinar archivos de cuadrículas múltiples.
- El marcado de puntos es relativamente más rápido.
- En Segmentación (Segmentation), los usuarios pueden visualizar diagrams de tipo MA y RI
- En Segmentación (Segmentation), los usuarios pueden elegir si sus criterios para el marcado automático serán combinados con operadores Boolean AND (todos) u OR (cualquiera).
- Datos "sin procesar" para todos los genes, incluyendo espacios o vacios, son ahora impresos en el archivo sin procesar, si el usuario decide crear un archivo sin procesar.
- En Explorar (Explore), los usuarios pueden crear diagramas de casilla para archivos y grupos de expresión para ver los resúmenes de 5 datos (valor mínimo, cuartil menor, media, cuartil mayor, y valor máximo).
- En Explorar (Explore), el usuario puede ahora encontrar genes mayores o menores de un cierto valor máximo, mínimo, o el valor del promedio mínimo absoluto.
- Los proyectos se cargan a una velocidad mucho más rápida que en la versión anterior.

#### Instalación de la Herramienta MAGIC

La Herramienta MAGIC es distribuida gratuitamente por Davidson College bajo la licencia pública GNU. Nuevas versiones de la Herramienta MAGIC pueden ser descargadas directamente yendo al sitio web de la Herramienta MAGIC:

# http://www.bio.davidson.edu/MAGIC

Empezando con la versión 1.5, la descarga de la Herramienta MAGIC consiste en un solo archivo de extensión Zip, llamado MAGIC\_Tool\_x-y.zip, el cual usted deberá descomprimir para ver la carpeta de la Herramienta MAGIC, llamado MAGIC\_Tool\_x-y. Los contenidos de la carpeta son descritos en la tabla siguiente:

Nombre del Archivo	Descripción
Magic_launch.bat	Archivo ejecutable para usuarios de Windows
MAGIC_launch	Archivo ejecutable para usuarios de MAC OS X
MagicTool.jar	Código de la Herramienta MAGIC (lanzado por el ejecutable)
MAGIC Users Guide v2-1.pdf	Guía del Usuario (en inglés)

Installation_guide.pdf	Instrucciones detalladas para instalar y ejecutar la Herramienta MAGIC
MAGIC Instructor's Guide.pdf	Guía del Instructor con detalles algorítmicos adicionales
Plugins	Archivos necesarios para TreeView de Java

Después de descomprimir el archivo descargado, diríjase hacia la carpeta MAGIC\_Tool\_x-y y haga doble clic en el archivo de ejecución apropiado para su sistema operativo (Windows o Mac OS X). Espere unos instantes, y la "pantalla splash" de la Herramienta MAGIC deberá aparecer, y en algunos instantes más el programa deberá abrirse. Si el archivo ejecutable no inicia apropiadamente el programa de la Herramienta MAGIC, vea la guía de instalación para instrucciones detalladas.

Archivos de muestra y el código de origen de la Herramienta MAGIC también son disponibles en el sitio web de la Herramienta MAGIC: <a href="http://www.bio.davidson.edu/MAGIC">http://www.bio.davidson.edu/MAGIC</a>

# Requerimientos del Sistema

- Windows 2000 o versión más actual O Mac OS X 10.4 o más actual O Unix/Linux
- JAVA JRE 1.5 (5.0) o versiones más actuales
- 512 MB de RAM requeridos para micro-colecciones de tamaño completo; no obstante, se recomienda tener un 1 GB de RAM en el sistema.
- Algunos cuantos cientos de MB (megabytes) de espacio en el disco duro disponibles, dependiendo de los archivos con los que usted trabaje y de los tipos de análisis que usted lleve a cabo.

#### Vocabulario

**Addressing** es el proceso corto de indicar a la Herramienta MAGIC la distribución de puntos y cuadrículas en el archivo tiff visto desde MAGIC.

Chip es un sinónimo para micro-colección.

**Puntos** es un sinónimo para un punto sencillo en una micro-colección.

**Flag** es un verbo que significa que usted puede marcar un punto en particular para indicar que sus datos no son fiables. Este puede ser el resultado de un fondo alto en el área, una pelusa o un grano de polvo sentado en el punto, etc. **Grid** es un arreglo compacto de puntos con espaciamientos uniformes.

**Gridding** es el proceso que la Herramienta MAGIC utiliza para encontrar puntos en sus archivos de extensión tiff.

**Metagrid** es un nivel de organización de orden alto. Una serie de cuadrículas son organizadas en grupos llamados "metagrids". Para una descripción más completa, por favor vea esta página web: <a href="www.bio.davidson.edu/projects/GCAT/Griding.html">www.bio.davidson.edu/projects/GCAT/Griding.html</a>

**Segmentation** es el proceso para encontrar la señal y distinguirla del fondo. Existen tres métodos en la Herramienta MAGIC. El "Fixed circle" (círculo fijo) es el más rápido, y es recomendado para la mayoría de los propósitos. "Adaptive circle" y "Seeded Region Growing" son también proporcionados como métodos. ["Adaptive circle" (Círculo adaptivo) y "Seeded Region Growing" (Crecimiento en Región Asemillada) son también proporcionados como métodos.)

Archivos **tiff** (por ejemplo nombre\_de\_archivo.tif) son los datos de imagen sin procesar que son producidos cuando una micro-colección de ADN es escaneada. Un archivo tiff es producido para cada color en cada chip.

#### Instrucciones Básicas

#### Información General de los Procedimientos

Si usted empieza con dos archivos tiff, usted necesitará ejecutar los siguientes procedimientos para producir grupos o explorar sus datos.

- 1. Comenzar la Herramienta MAGIC
- 2. Comenzar un proyecto
- 3. Agregar archivos al proyecto (recomendado)
- 4. Cargar archivos tiff
- 5. Cargar la lista de genes
- 6. Localizar puntos (Gridding y Addressing)
- 7. Distinguir la señal del fondo y generar archivos de expresión (Segmentation)

- 8. Repetir los pasos 1-6 para todas las condiciones experimentales, añadiendo los datos anteriores y formando un archivo de expresión con varias columnas.
- 9. Proporciones de log-transform (transformación por logaritmos)
- 10. Agregar información de genes al archivo de expresión (opcional y en inglés)
- 11. Explorar datos (recomendado)
- 12. Filtrar datos (recomendado)

Los siguientes procesos pueden ser ejecutados solamente si usted tiene tres o más columnas en su archivo de expresión:

- 13. Calcular la disimilaría (por ejemplo una 1- correlación)
- 14. Agrupar genes

#### (1) Comenzar un Proyecto

En el menú "Project" (Proyectos), elija "create a new Project" (crear un nuevo Proyecto). Usted puede guardar este proyecto en un directorio conveniente en su disco duro. Nosotros recomendamos NO utilizar la carpeta del software MAGIC Tool, ya que usted podría desear en el futuro abrir este proyecto con una nueva versión de la Herramienta MAGIC. A los archivos del proyecto se les asigna automáticamente un nombre que termina con el sufijo ".gprj" y son almacenados en una carpeta que contiene el mismo nombre, creado de forma automática por la Herramienta MAGIC.

#### (2) Agregar Archivos al Proyecto

Nosotros recomendamos que usted copie los archivos en su proyecto, ya sea utilizando las opciones del menú "Project" (Proyectos), o arrastrando los archivos hacia la carpeta del proyecto en cuestión y después seleccionar el comando "Update Project" (Actualizar Proyecto) del menú "Project". Agregar archivos a su proyecto es un proceso que le organiza sus archivos en fólderes predeterminados, y simplifica procesos futuros al momento de efectuar un análisis.

# (3) Cargar Archivos Tiff (Control R y Control G)

En el menú "Build Expression File" (Construir Archivo de Expresión), cargue los pares de imagenes rojo y verde. Recuerde que el rojo es una longitud de onda más larga que la verde,

Load Image Pair...

▶ □ Red: <none> Ctrl-R

Green: <none> Ctrl-G

de manera que sus archivos son identificados por las longitudes de onda, y por lo tanto usted aún puede determinar cuál color es cuál.

#### (4) Cargar Lista de Genes (Control X)

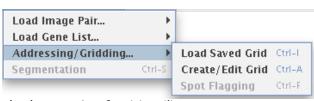
Cargue la lista de genes; ejecute este comando también usando el menú "Build Expression File (Construir Archivo de Expresión). Éste deberá ser



un archivo de texto con un sufijo ".txt" y deberá estar en un formato de la Herramienta MAGIC. (Vea por favor las instrucciones abajo.)

# (5) Localizar Puntos

En el menú "Build Expression File" (Construir Archivo de Expresión), seleccione la opción "Addressing/Gridding". Hay varios pasos diferentes en Addressing y en Gridding, de los cueles hablarames una par una en la



de los cuales hablaremos uno por uno en los siguientes párrafos (a) -(j).

# (a) Decida si usted quiere crear una cuadrícula nueva o si quiere cargar una cuadrícula anterior.

A menos que usted haya hecho esto antes, usted deberá crear una nueva cuadrícula. Si usted tiene una cuadrícula previamente creada que es apropiada para esta imagen, usted puede cargarla simplemente al escoger la opción "Load Saved Grid" (Cargar Cuadrícula anterior) localizada en el submenú Addressing/Gridding, o al presionar la combinación de teclas "Control + W", y proceder directamente al paso 6, Segmentation (Segmentación). Para crear una nueva cuadrícula, seleccione "Create/Edit Grid" (Crear o Correjir Cuadrícula) del submenú Addressing/Gridding, localizado en el diagrama de la parte posterior, o presione la combinación de teclas Control + A.

Cuando usted crea una cuadrícula nueva, aparecerá una ventana de advertencia que es normal e intencional. La advertencia es un recordatorio que usted DEBE comprender la forma en la cual sus puntos están arreglados en su micro-colección. Para más información acerca de este paso, consulte

http://gcat.davidson.edu/GCAT/workshop2/addressing MT.html



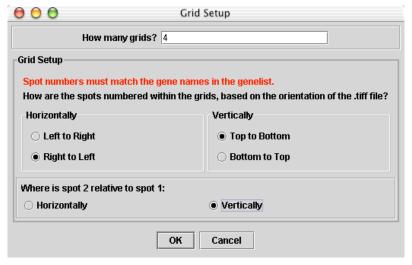
No proceda más allá de este paso si usted no comprende la organización de su micro-colección.

El no ejecutar correctamente los procesos de Addressing y Gridding resultará en una identificación incorrecta en los puntos de su micro-colección.

Usted deberá ver dos ventanas. La primera le mostrará sus archivos tiff mezclados y la segunda le permitirá efectuar el proceso de "Addressing" a su archivo tiff. La ventana más pequeña (desplazable) le preguntará información acerca de la organización de su microcolección, a esto se le llama "Addressing".

#### (b) Conteste las cuatro preguntas en la ventana Grid Setup (Montaje de Cuadrícula)

Primero, ingrese el número total de cuadrículas en el archivo tiff.

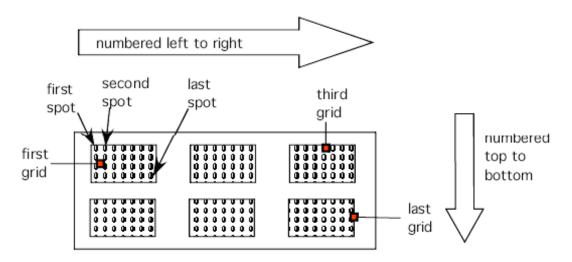


Las respuestas a las otras tres preguntas es el paso más fácil para cometer un error desastroso. Conteste las tres preguntas basándose en la forma en la que usted está viendo su microcolección en este momento. Aquí hay un ejemplo para demostrar este punto. Suponga que la imagen ha

sido rotada a un ángulo de 90 grados en sentido de las manecillas del reloj comparada a la forma en la cual usted normalmente considera su chip, pero su lista de genes no está alterada para tomar en cuenta la rotación. Entonces la forma en la que usted está viendo su archivo tiff no será igual a la concepción mental que usted tiene de la organización de su micro-

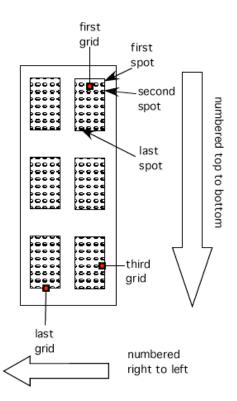
colección. Las siguientes dos imágenes muestran el diseño de la micro-colección antes y después de la rotación.

Antes de la rotación, los puntos serían descritos como una enumeración de arriba hacia abajo y de la izquierda a la derecha, con el segundo punto horizontal con respecto al primer punto (de la misma forma en la que usted leería un libro). Éstas son las opciones predeterminadas. Sin embargo, es importante que usted lleve la cuenta de los puntos si el chip es rotado.



Después de la rotación, los puntos son numerados de arriba hacia abajo, de la *derecha* hacia la *izquierda*, y el segundo punto está ahora *vertical* en relación al primer punto (abajo). Estudie las imágenes que describen la organización antes y después de la rotación para entender cómo los puntos han sido movidos y porqué la nueva orientación resultó en el "addressing" que se proporcionó en el ejemplo de arriba. Ahora estudie todas las otras opciones para enumerar los puntos en la tabla de abajo.

Si es necesario, utilice el patrón de puntos ausentes y los comentarios en su lista de genes para que usted pueda reorientarse. También, la distribución y el



número de cuadrículas son maneras fáciles de orientarse.

Si usted comete un error, usted puede cambiar sus respuestas a estos problemas de "addressing" al seleccionar el comando "Grid properties..." (Propiedades de la cuadrícula) en el "File menu" (Menú del Archivo) de la ventana de la cuadrícula.

	Horizontally L	EFT to RIGHT	
Vertically TO	OP to BOTTOM	Vertically BO	OTTOM to TOP
Spot 2 Horiz of Spot 1	Spot 2 Vertical of Spot 1	Spot 2 Horiz of Spot 1	Spot 2 Vertical of Spot 1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	1 8 15 2 9 16 3 10 17 4 11 18 5 12 19 6 13 20 7 14 21	21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1	7 14 21 6 13 20 5 12 19 4 11 18 3 10 17 2 9 16 1 8 15
		IGHT to LEFT	
	OP to BOTTOM		OTTOM to TOP
Spot 2 Horiz of Spot 1	Spot 2 Vertical of Spot 1	Spot 2 Horiz of Spot 1	Spot 2 Vertical of Spot 1
3 2 1 6 5 4 9 8 7 12 11 10 15 14 13 18 17 16 21 20 19	15 8 1 16 9 2 17 10 3 18 11 4 19 12 5 20 13 6 21 14 7	21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1	21 14 7 20 13 6 19 12 5 18 11 4 17 10 3 16 9 2 15 8 1

# (c) Empiece a cuadricular.

El propósito de cuadricular es aquél que indica a MAGIC dónde se localizan los puntos en cada cuadrícula. Esta función es una de las mejores innovaciones en la Herramienta MAGIC. Antes de empezar, usted querrá ajustar el contraste para ayudar ilustrar los puntos que están poco sombríos. Para hacer esto, deslice el indicador que está actualmente apuntando a contraste 100% cerca de la parte superior de la ventana. Ajustar el contraste no afecta los datos sin procesar; tan sólo le permite ver los puntos de una forma más clara en este paso.

La primera lengüeta o etiqueta (tab) deberá ser seleccionada como el "default" (punto de referencia) cuando usted empiece el proceso de gridding. Los números de las etiquetas en la micro-colección corresponden a los números en la cuadrícula. Seleccionar la etiqueta #1 indica que usted está trabajando con la cuadrícula #1 (basada en la lista de orden de genes). Usted puede empezar con una matrícula diferente si así lo desea, pero asegúrese de saber dónde se encuentra cada cuadrícula en la micro-colección. Le recordamos nuevamente que si usted no sigue este procedimiento de emparejar los números de la cuadrícula con los números de la etiqueta, usted provocará que los "puntos" (funciones) sean identificadas

incorrectamente. La cuadrícula #1 es la cuadrícula que contiene el punto #1, que corresponde al gen #1 en la lista de genes.

#### (d) Centre la cuadrícula actual en la ventana de cuadrículas.

Desplace y magnifique la imagen hasta que usted pueda ver la primera cuadrícula definida por la lista de genes. Para hacer una mayor magnificación, haga clic en el botón "Zoom In" y después haga clic en la cuadrícula donde usted quiera que el acercamiento se centre. Recuerde que los puntos y los genes no cambian sus números con la rotación de imágenes. En el ejemplo de arriba donde la imagen es rotada 90 grados en sentido de las manecillas del reloj, la primera cuadrícula sería la cuadrícula en la esquina

derecha superior.

# (e) Entre la localización de la cuadrícula usando el método "3click" del ratón.

- Haga clic en el botón que dice "Set Top Left Spot" y después haga clic en el centro del punto de la izquierda superior en la cuadrícula.
- Haga clic en el botón que dice "Set Top Right Spot" y después haga clic en el centro del punto de la derecha superior en la cuadrícula.
- c. Haga clic en el botón que dice "Set Bottom Row" y después haga clic en el centro de cualquier punto en la fila inferior. Elija un punto redondo grande para hacer este paso más fácil.
- d. Ingrese el número de filas y columnas. Esto será contestado basándose en la forma en la que usted está actualmente viendo el archivo tiff. En este ejemplo, hay 24 filas y 12 columnas.
- e. Haga clic en el botón "Update" (Actualizar). En este momento, usted puede deslizar el cursor sobre cualquier punto e identificar su localización (coordenadas dadas en X y Y en pixeles, fila, columna, y número de punto) así como su identidad obtenida de la lista de genes. Esta información aparece en la esquina izquierda inferior y es especialmente útil para navegar durante el proceso de "Segmentation" (Segmentación).

X:133 Y:353 Gene:YMR186W (Grid:1 Col:7 Row:18 Spot Number:162)

# (f) Ajuste la cuadrícula para centrar los puntos.

El margen desde el siquiente parafo hasta terminar con el paso (6) (página 19) debe moverse a la izquierda, chequee (d) como ejemplo (Indent)

En este momento, vea si los puntos se ven centrados en las casillas. Si éste no es el caso, entonces ajuste la posición de las casillas ya sea haciendo clic en el botón apropiado y después en el punto correcto, escribiendo los números correctos para ajustar las casillas, o ajustando la cuadrícula con el ratón. Si usted hace clic en cualquier parte del interior de la cuadrícula, usted puede arrastrar toda la cuadrícula a una nueva localización. La cuadrícula puede ser reajustada en tamaño desde una esquina al hacer clic en uno de los puntos grises y arrastrando el ratón. Mientras usted hace esto, simultaneamente se indica el nuevo tamaño y la posición de la cuadrícula. Finalmente, si usted hace clic en uno de los botones de rotación, toda la cuadrícula va rotar alrededor de su centro, permitiéndole ajustar cuadrículas un tanto inclinadas en sus imágenes. Si usted decide manualmente ajustar la cuadrícula al cambiar los valores en las casillas, tenga en cuenta que la posición del ratón se indica en la esquina izquierda de la ventana de forma que usted puede determinar si los números deberían ser más grandes o más pequeños para ajustar las casillas en la dirección correcta. Hacer el proceso de cuadricular (Gridding) toma un poco de práctica, pero es MUCHO más fácil que la mayoría de los métodos de Gridding.

#### (g) Defina la siguiente cuadrícula.

Si usted tiene sólo una cuadrícula, pase al paso (i). Si usted tiene más de una cuadrícula, continúe leyendo. Una vez que la primera cuadrícula ha sido propiamente rodeada con las casillas con los puntos en los centros, es tiempo de repetir este proceso para la cuadrícula #2. Asegúrese de saber si la cuadrícula #2 se encuentra a la izquierda, a la derecha, arriba o debajo de la cuadrícula #1.

Presione y mantenga la tecla Control (Ctrl) en el teclado, y después haga clic en el centro del punto de la esquina superior izquierda de la cuadrícula #2. La misma cuadrícula traducida a la localización especificada por el clic del ratón aparecerá como la cuadrícula #2, y todos los números en las casillas de la izquierda serán llenados automáticamente. Si usted deja de presionar la tecla Control, usted puede ajustar la cuadrícula como usted lo hizo previamente en el paso *f*. Repita este proceso para todas las cuadrículas.

#### (h) Continúe el proceso de cuadricular.

Continúe el paso (g) para cada cuadrícula restante en la micro-colección, de forma que todas las cuadrículas en la micro-colección sean encasilladas con los puntos en el centro de las casillas. En cualquier momento, usted puede cambiar sus respuestas a los cuatro problemas de "Addressing" al seleccionar las "Grid Properties..." (Propiedades de la Cuadrícula) en el "file menu" (menú del archivo" de la ventana de la cuadrícula.

Si usted necesita mover múltiples cuadrículas a la vez, presione y mantenga la tecla Shift, y después haga clic en cada cuadrícula que usted desee mover. Al seleccionar las cuadrículas, éstas serán resaltadas en color azul. Una vez que todas las cuadrículas que

usted quiera mover estén visualizadas en azul, haga clic y arrastre adentro cualquiera de las cuadrículas para moverlas todas a la vez. Usted también puede rotar simultáneamente múltiples cuadrículas al seleccionarlas de la misma forma y hacer clic en uno de los botones de rotación.

Usted puede parar en cualquier momento y guardar el progreso de su trabajo, utilizando el submenú "Save Current Grid As..." (Guardar Cuadrícula Actual Como...) en el "file menu" (menú del archivo) de la ventana de la cuadrícula. La próxima vez que usted empiece el proceso de Addressing/Gridding, usted puede simplemente abrir este archivo de cuadrícula guardado.

Si usted crea dos diferentes archivos de cuadrícula, usted los puede combinar usando la opción "Combine and Load Grid Files" (Combinar y Cargar Archivos de Cuadrícula) en el menu "Build Expression File" (Construir Archivo de Expresión). Cuando usted elija esta opción, se le pedirá seleccionar el primer archivo de cuadrícula. De este archivo, la Herramienta MAGIC tomará los detalles de orientación de la cuadrícula que usted previamente determinó en el paso (b), además de tomar todas las cuadrículas en este archivo. Una vez que usted seleccione el primer archivo de cuadrícula, se le pedirá seleccionar el segundo archivo de cuadrícula, y después el nuevo nombre de archivo para el archivo de cuadrícula combinado. La Herramienta MAGIC toma las cuadrículas del primer archivo como las primeras cuadrículas n en el nuevo archivo seguido por las cuadrículas del segundo archivo como el resto de las cuadrículas. Usted deberá asegurarse que las cuadrículas sean combinadas en el orden correcto. Una vez que el archivo de cuadrículas sea creado, la Herramienta MAGIC cargará e indicará el archivo de cuadrícula combinado, y usted podrá editar la cuadrícula al seleccionar la opción "Create/Edit Grid" (Crear/Editar Cuadrícula), o continuar directamente al paso 6 "Segmentation" (Segmentación).

Usted también puede guardar una foto de la pantalla de las imágenes combinadas de extensión tiff en cualquier momento antes del o durante el proceso de cuadricular (Gridding). Usted puede guardar la imagen con un formato tiff, jpg o gif. El formato tiff funciona en todos los procesadores de texto y de dibujo lo que lo hace un formato universal. El formato jpeg es una buena opción para imágenes como éstas que pueden contener muchas sombras, de la misma forma que uno utilizaría el formato jpeg para una fotografía. El formato gif es el más simple, lo cual puede hacer que usted pierda la sutileza en detalles original de su archivo. Esta imagen mezclada guardada es útil si usted quiere tomar una fotografía de toda la cuadrícula, la cual puede utilizar para publicaciones o fines didácticos.

#### (i) Complete el proceso de cuadricular.

Cuando usted haya acabado el proceso de "Gridding" en todas sus matrículas, haga clic en el botón "Done!" (¡Terminado!). Si usted no ha guardado todavía su cuadrícula, el programa le pedirá hacerlo antes de poder avanzar al siguiente paso. Un archivo de cuadrícula deberá ser guardado en la carpeta de su proyecto y le será automáticamente asignado un sufijo de ".grid" (de manera que usted no tiene que escribir ".grid")

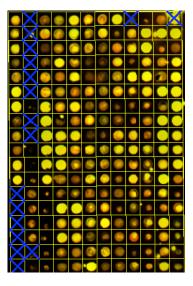
Si el número de genes en su lista de genes y el número de puntos que usted cuadriculó no no son iguales, usted obtendrá un mensaje de error. Usted debe tener exactamente un cuadrado de la cuadrícula por cada línea (gen o replicado de gen) en la lista de genes. Si éste no es el caso, usted cometerá probablemente un error identificando los puntos más adelante durante el análisis, de forma que se le pide corregir el problema en este paso. Si la lista de sus genes y el número de puntos cuadriculados son iguales, entonces usted será informado del número total de puntos y le será permitido guardar el archivo de cuadrícula para uso posterior.

# (j) Determinar los puntos problemáticos (opcional)

Si hay puntos en su cuadrícula que usted no desea que sean utilizados en su análisis de

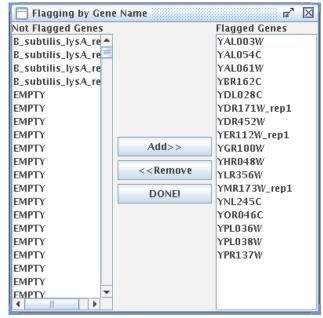
datos, usted puede escoger excluir los datos en este momento, antes de la creación del archivo de expresión. Para hacer esto, elija "Spot Flagging" del submenú "Addressing/Gridding", o presione Control + F.

De la misma forma que en la ventana gridding, usted puede hacer un zoom in (acercamiento) o un zoom out (acercamiento), y ajustar la imagen a la pantalla. Y también como lo hizo en la ventana de la cuadrícula, cuando usted desplaza el ratón sobre un punto, la barra de estado en la parte inferior de la ventana le mostrará la información acerca del gen. Si usted ve un punto que usted no quiere sea incluido en sus cálculos, haga clic sobre dicho punto. Una cruz "X" de color azul aparecerá sobre el punto marcándolo como "flagged" y el proceso de segmentación lo ignorará.



Para ver cuáles genes han sido "flagged", o para elegir los genes que serán marcados o aquéllos que no deberán serlo en función de su nombre, seleccione "Flagging by Gene Name" del menú Flagging. En el ventana de diálogo que aparece, los genes no marcados (aquéllos que serán usados) aparecen en la izquierda de la pantalla, y aquéllos que son marcados aparecen en la derecha. Para marcar un gen, haga clic (y por ende resaltándolo en la lista) de la izquierda y haga clic en "Add" (Agregar). Para desmarcar un gen, haga clic en su nombre en la lista de la derecha y después haga clic en "Remove" (Quitar). Usted puede seleccionar múltiples genes en la lista al mantener presionada la tecla

"Control", y después al hacer clic en cada gen, o, para seleccionar un rango de genes, haga clic en el primero, después mantenga presionada la tecla "Shift", y después haga clic en el último gen. Una vez que usted haya presionado el botón "Add" o "Remove" (Agregar o Quitar), los cambios se vuelven visibles en la imagen detrás de la ventana "Flagging by Gene Name" (marcador por nombre de gen). Note que los genes con nombres "empty", "missing", "none", o "blank" (en español: vacío, ausentes, ninguno, o blanco) son

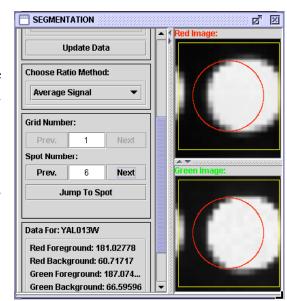


automáticamente excluidos del archivo de expresión, así que no necesitan ser marcados por nombre. Cuando usted haya acabado de marcar genes por nombre del gen, haga clic en "Done!" (¡Listo!).

De la ventana "Flagging" principal (Marcación), usted también puede ya sea cargar o guardar archivos de marcación. Estos archivos tiene la extensión ".flag" y se encuentran guardados en la subcarpeta titulada "flags" en la carpeta principal que contiene su proyecto. Este proceso de guardar su trabajo funciona de la misma forma que aquél descrito para el cuadriculado, cuyos pasos se encuentran detallados en los párrafos (h) e (i) arriba, excepto que en este proceso a usted no se le pide guardar un archivo de marcación automáticamente. Para cargar un archivo de marcación, abra la ventana "Spot Flagging" (Marcación de Puntos) y después elija "Load Saved Flags" (Cargar

Marcaciones Guardadas...) del menú File (Archivo). En esa ventana, usted puede escoger el archivo de marcación que usted quiera cargar. Note que el número de cuadrículas y el número de puntos por cuadrícula tiene que ser el mismo de la cuadrícula actual para que usted pueda cargar el archivo de marcación.

(6) Distinga una señal de fondo y genere un archivo de expresión (Segmentación; Control + S)

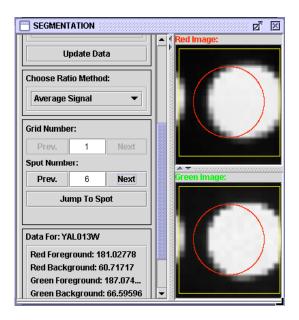


A continuación se divide este paso en tres partes, descritas en los párrafos (a) – (d)

# (a) Seleccione un método para distinguir qué es señal y qué es fondo.

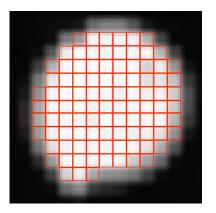
Círculo fijo: La manera más común de hacer esto es simplemente poner un círculo en medio de los cuadros que usted dibujó para el proceso de cuadriculación. A esto se le llama un *circulo fijo*, aunque usted puede ajustar el radio de dicho círculo como se muestra en la figura en la derecha. Note que aún si el círculo es más grande que la casilla, solamente la señal dentro de la casilla es usada para medir la señal.

**Círculo adaptivo:** El segundo método es el *círculo adaptivo*. El tamaño y la localización del círculo cambia, dependiendo del tamaño en los puntos en la micro-colección. Vea la Guía del Instructor para más detalles en este algoritmo.



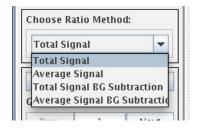
Crecimiento en Región Asemillada: El crecimiento en región asemillada está diseñado para encontrar la señal en cada punto basado en la distribución de la señal. Este método

de segmentación busca el pixel más brillante cerca del centro de la cuadrícula, y después conecta todos los pixeles adyacentes a dicho pixel y los conecta en una forma. El algoritmo conecta de forma simultánea los pixeles al fondo y a la región de primer plano, continuando hasta que todos los pixeles estén en una de las regiones. Un límite especificado por el usuario determina cuáles pixeles pueden ser utilizados para "asemillar" la región. Éste es el método más lento ya que cada pixel es procesado individualmente. En general, mientras más grande el límite sea, el punto será definido de una forma más grande.



# (b) Elija un Método de Proporción

El producto final de la segmentación es una lista de proporciones de expresiones de genes. Hay cuatro opciones para elegir cómo combinar los cuatro valores numéricos en la segmentación (rojo en primer plano, rojo en el fondo, verde en primer plano, verde en el fondo) para determinar una proporción para cada punto en la



micro-colección. La opción "total signal" (señal total) agrega los valores en todos los pixeles diseñados como señal, y divide el total de rojos entre el total de verdes. El comando "average signal" (señal promedio) computa el promedio de los valores de los pixeles de señal. Las dos opciones restantes substraen el fondo (total o promedio, respectivamente) antes de dividir el rojo entre el verde para obtener la proporción. La substracción del fondo permite la posibilidad de un valor negativo (si el fondo es más grande que el primer plano). La herramienta MAGIC transforma un valor negativo a 0. Si el fondo es más grande o igual al primer plano en la señal verde, esto resulta en división por 0. En este caso, la herramienta MAGIC restablece la proporción a 998 o 999) (dependiendo en si el numerador del primer plano rojo menos el fondo rojo también era 0, o si era más grande que 0).

Usted puede navegar alrededor de los puntos, mirando el resumen de los datos de cada punto abajo, para verificar visualmente que la cuadriculación y la segmentación han sido adecuadamente ejecutadas. Esta inspección le da la oportunidad de chequear cualquier punto que usted piensa no debería ser considerada durante próximos análisis de datos.

#### (c) Marcar Puntos Automáticamente (opcional)

Una vez que usted ha escogido el método de segmentación y el método de proporción, usted puede determinar los criterios de forma como por ejemplo, si cualquier punto no cumple con las normas, su proporción no será incluida en el archivo de expresión. Para hacer esto, haga clic en el botón "Automatic Flagging Options" (Opciones de Marcado Automático). Aquí, usted puede determinar valores límites para estos criterios, y escoger marcar un punto ya sea si uno (operador Boolean OR) o todos (operador Boolean ALL) los requisitos han



sido cumplidos para dicho punto. Cuando usted haga clic en OK, se le preguntará si usted desea ejecutar cálculos o no para encontrar el nivel de marcación de los puntos. En el proceso, la Herramienta MAGIC también computa el promedio y la deviación estándar para cada uno de los cuatro puntos de información utilizados en los cálculos (aun si usted ha dejado los límites en blanco). Entonces, usted puede utilizar estos datos para refinar los requisitos de marcado automático. Por ejemplo, si es que usted desea marcar aquellos genes cuyo primer plano total rojo o verde son menos de dos deviaciones estándar debajo del promedio.

Para ver en una cuadrícula cuáles puntos han sido marcados, abra la ventana "Spot Flagging" del submenú "Addressing/Gridding". Todos los puntos que han sido automáticamente marcados serán marcados con una "X" naranja. Estas banderas pueden ser únicamente cambiadas al ajustar los requisitos de marcado automáticos, pero también usted puede agregar o eliminar marcaciones automáticas en este momento. Si un punto es marcado de forma automática y manual, una "X" azul aparecerá superimpuesta en el punto en lugar

Summary Statistics

Red FG Average: 2005775.9436

Red FG Std. Dev.: 1556768.4052

Red BG Average: 1343939.1146

Red BG Std. Dev.: 1047404.7682

Green FG Average: 1640906.6979

Green FG Std. Dev.: 1469862.3653

Green BG Average: 1046154.8307

Green BG Std. Dev.: 948306.8859

de la "X" naranja. Si usted desmarca de forma manual un punto que ha sido previamente marcado automáticamente, la "X" se volverá de color naranja y el punto seguirá siendo contado como marcado.

Si usted ajusta las opciones de marcado automático, usted deberá calcular nuevamente los datos para obtener los marcados automáticos y hacer que éstos aparezcan en la pantalla "Spot Flagging" (Marcación de Puntos). Cuando usted haya acabado de ajustar las opciones a su satisfacción, continúe a generar el archivo de expresión.

#### (d) Generar archivos de expresión

Haga clic en "Create Expression File" (Generar Archivos de Expresión) cuando usted esté satisfecho/a con el proceso de segmentación. Esto generará un archivo de expresión, el cual ha sido el propósito de todos los pasos anteriores. Un archivo de expresión contiene las proporciones para cada punto (rojo dividido entre verde), según el método elegido. La herramienta MAGIC ignorará algunos registros en la columna que contiene los nombres de los genes ("blank", "EMPTY", "missing" y "none"; (valores que significan "sin respuesta" "VACIO" "ausente" y "ninguno") no importa si los valores están escritos en minúsculas o mayúsculas). Las proporciones serán utilizadas en todos los subsecuentes

análisis de datos. Usted ya no necesita los archivos tiff.

A menos que usted ya haya creado un archivo de expresión para esta micro-colección, usted deberá seleccionar la casilla junto a la pregunta "Create

✓ Create Expression File?					
Enter Expression Filename: cell_cycle					
Enter Colun	ın Name: t0				
Create !	New File				
<ul><li>Append</li></ul>	To File: Please Select A File				
1 I D.	Name (Gene Names Must Match Exactly)				
Append By:	O Name (List Orders Must Match Exactly)				

Expression File?" (¿Crear Archivo de Expresión?), y deberá nombrar tanto al archivo de expressión como a la columna (esto es período de tiempo, tratamiento, etc.). Usted puede anexar esta columna a un archivo existente o crear un nuevo archivo de expresión que contenga únicamente esta columna. La herramienta MAGIC nunca borrará ningún archivo, así que si usted anexa esta columna a un archivo de expresión existente, ese archivo permanecerá como estaba en su computadora, y un nuevo archivo será creado con la columna existente anexada a la derecha de las columnas en el archivo en cuestión.

En la casilla de diálogo "Expression File Parameters" (Parámetros del Archivo de Expresión), usted también puede escoger si usted quiere guardar los datos "sin procesar" que fueron utilizados para computar las proporciones de expresión. Si usted selecciona la casilla que se encuentra junto a la opción "Create Raw Data File", un archivo de texto limitado por etiquetas será creado y contendrá 9 columnas. La primera columna corresponde al nombre del gen. Las siguientes cuatro columnas son los totales de los pixeles para el primer plano de color rojo, el fondo de color rojo, el primer plano verde, y el fondo verde. Las últimas cuatro columnas son los promedios de los pixeles para estos mismos cuatro valores. El archivo de datos sin procesar tendrá el mismo nombre que la etiqueta de su columna, con la extensión ".raw". Su computadora podría tomar el archivo como un archivo de imagen, pero es sólo un archivo de texto. Usted puede abrir este archivo desde Excel (usted tendrá tal vez que activar la opción Buscar todos tipos de archivos para que el programa puede abrir el archivo). En futuras versiones de la Herramienta MAGIC, usted podrá utilizar los datos sin procesar para filtrar sus datos de expresión, por ejemplo cuando las señales sean demasiado débiles para que puedan ser consideradas como fiables. Mientras tanto, este tipo de filtración debe ser hecho en una aplicación externa al programa de la herramienta MAGIC.

# (7) Repita los pasos 1-6 para todas las condiciones experimentales

Si usted tiene múltiples períodos de tiempo o condiciones experimentales en su estudio, usted deberá repetir los pasos 1-6 para cada condición antes de continuar a las manipulaciones de datos del paso 8. Una vez que usted tenga todos los datos en un archivo, continúe con los pasos restantes. Si usted tiene solamente una condición, sólo habrá una columna de datos en su archivo, y usted podrá ejecutar los pasos 8-11.

#### (8) Manipule Datos

Aunque este paso pueda parecer un intento de conducir fraude científico, es de hecho un paso beneficial a considerar (vea la Guía del Instructor). Usted puede: transformar o normalizar sus datos; temporalmente restringir su análisis de datos a un subgrupo de condiciones experimentales (por ejemplo a ciertos períodos de tiempo o intercambio de tintura); filtrar

algunos puntos que no cumplan conciertos requisitos; o generar un grupo aleatorio de datos para usar como punto de comparación.



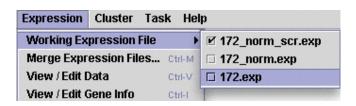
Si usted manipula sus datos, usted generará una colección nueva de archivos de expresión con nombres que sean iguales a los de la manipulación. La Herramienta MAGIC nunca borrará sus datos, de forma que el resultado de cualquiera de estas manipulaciones de datos es guardado en un archivo nuevo, y el archivo original existirá todavía como solía ser antes de la

manipulación. Esté seguro de verificar cuál es el archivo de expresión con el que usted está trabajando para los pasos subsecuentes. Es muy fácil confundirse. El archivo actual tiene un signo visto en la lista "Working Expression File" (Archivo de Expresión de Trabajo).

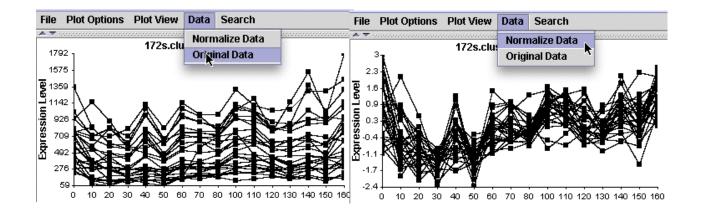
Si usted está trabajando con datos de proporción, usted deberá transformar sus datos de forma "log" (logaritmo). Esto convertirá sus proporciones en valores que están en la misma escala numérica de forma que un gen que es inducido cuatro veces (+2) tiene el mismo valor numérico que un gen reprimido que tiene un "4 fold" (-2 en lugar de 0.25). Generalmente, esto se hace empleando una transformación "log2", la cual indica el número de cambios "two fold" en una expresión de genes (de ahí los cambios 4 fold resultaron en valores numéricos de 2).

Si usted está trabajando con valores de expresión absoluta (por ejemplo, datos Affymetrix) a usted le convendría normalizar sus proporciones.

Normalizar en este caso, se hace



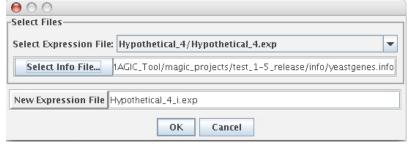
pensado de gen en gen. Para cada gen, el valor medio a través de las columnas se sustrae de cada valor, resultando en un perfil de expresión con una media de 0. Después cada valor es dividido entre la deviación estándar a través de las columnas, resultando en un perfil de expresión con una deviación estándar de 1. Este tipo de normalización es especialmente útil para ver grupos de genes en la misma escala, así que las similitudes son más fácilmente vistas cuando los niveles de expresión absolutos varían considerablemente de gen a gen. Después, cuando usted grafique los varios grupos o las agrupaciones de genes, usted podrá ver los datos en valores ya sea normalizados u originales, como se demuestra en la siguiente figura.



# (9) Agregar información de genes al archivo de expresión

Ahora es el tiempo predilecto para agregar anotaciones de genes a su archivo de expresión, para

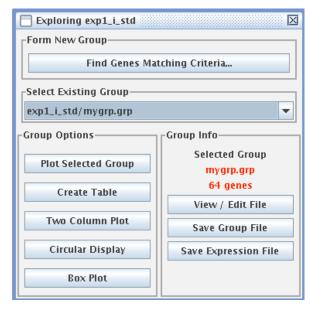
que dichas anotaciones y comentarios sean visibles al momento de explorar sus datos. En el menú "Expression" (Expresión), escoja "Import Gene Info..." (Importar Información de Gen). Seleccione el archivo de



expresión al que usted quiera agregar las anotaciones, y seleccione el archivo que contenga dichas anotaciones. Un archivo que contiene anotaciones de levaduras está incluido en los archivos de ejemplo. Un archivo similar puede ser formado para cualquier organismo al crear un archivo de texto delimitado por etiquetas con las columnas apropiadas (alias, cromosoma, localización en el cromosoma, proceso biológico, función molecular, y componente celular).

# (10) Explorar datos

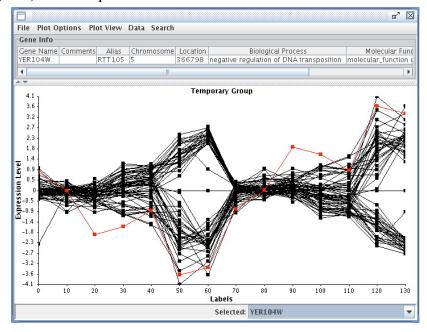
La exploración de datos es una manera de familiarizarse con sus datos, y encontrar relaciones funcionales importantes que tal vez no sean aparentes al momento de conducir la agrupación. Por ejemplo, usted puede encontrar todos los genes que incrementaron su expresión después de un cierto período de tiempo, o todos los genes que incrementaron su represión (disminuyeron su expresión) de pliegue cuatro veces o más en cualquier período de tiempo. Una



vez que usted haya identificado dichos genes, usted los puede mostrar en una variedad de formas

dinámicas y guardar estas imágenes para publicaciones futuras o con fines didácticos.

Si usted no ha explorado el archivo de expresión actual hasta ahora y si no ha guardado los archivos de grupo, el único grupo existente es el archivo completo de expresión. Usted puede crear un grupo temporal al hacer clic en "Find Genes Matching Criteria..." y llenar la información para encontrar los genes y los patrones de expresión que a usted le interesan. Si usted quiere que un grupo este disponible la próxima vez que usted explore sus datos, la próxima



vez que usted abra este proyecto, usted deberá guardar el archivo de grupo (el cual será asignado una extensión automática de ".grp"). Un archivo de grupo es simplemente un archivo de texto que lista los nombres de los genes en el grupo. Cualquier grupo guardado será enlistado en "Select Existing Group" (Seleccione Grupo Existente). Un grupo de genes puede ser guardado también como un archivo de expresión, el cual guarda todas las columnas de proporciones o las proporciones logarítmicas junto con los nombres de genes.

Cada uno de las pantallas del lado izquierdo de la ventana exploradora le ofrece una visualización diferente de sus datos. La pantalla "Plot Selected Group" (Grupo Seleccionado de Diagramas) se ilustra aquí, con el gen YER104W resaltado. Note que los comentarios de este gen pueden ser revelados arriba del diagrama del grupo. Este grupo ha sido formado al encontrar todos los genes cuyo valor mínimo ha sido menor a -2. Cabe resaltar que este grupo parece consistir de dos sub-grupos diferentes: genes que son incrementados de forma temporal y disminuidos después, y genes que tienen el perfil opuesto.

#### (11) Filtrar datos

Usted debe filtrar sus datos para eliminar genes que no le interesen antes de proceder a los próximos pasos. Por ejemplo, usted tal vez mantenga solamente los genes cuyo patrón de expresión tenga una deviación estándar lo suficientemente larga a través de las columnas (en otras palabras, cuya expresión no sea constante). O usted puede eliminar genes con proporciones no fiables (incluyendo aquellos que involucren una división entre 0). Es importante que su archivo de expresión sea lo más pequeño posible, sin perder información importante, antes de

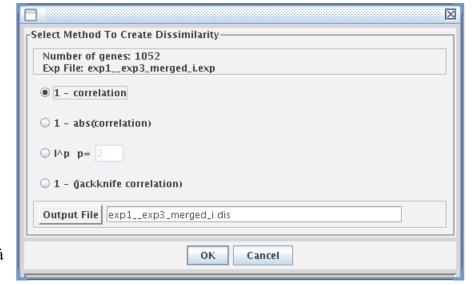
empezar el proceso de agrupamiento. Usted puede filtrar al guardar los resultados de búsquedas en la ventana "Exploring" como un archivo de expresión.

# (12) Calcular disimilitudes

Para formar agrupamientos de genes similares usted necesita una manera de comparar los perfiles de expresión de genes diferentes. En este paso, usted generará una tabla enorme en "disimilitudes", midiendo la diferencia entre cada patrón de expresión de genes. Este paso puede tomar muchísimo tiempo si utiliza un gran número de genes. Asegúrese de que usted ha filtrado sus datos de forma adecuada, y que usted sepa que usted aprenderá algo del proceso de

agrupamiento antes de comenzar este paso.

En el menú "Expression" (Expresión), escoja "Dissimilarities" (Disimilitudes) y después "Compute" (Computar). Cuando usted haga esto, una ventana aparecerá donde usted tendrá tres opciones para escoger. Ésta es otra decisión que afectará el análisis de datos.



El método más común es el predeterminado, que es la correlación-1. El segundo método, abs(correlación)-1 |correlación|-1, es similar al método de correlación-1, pero el valor absoluto del coeficiente de correlación es tomado antes de que dicho número sea substraído del 1. ||

Este método le da una medida de qué tan cerca los genes relacionados aparentan estar sin tomar en cuenta si la correlación es positiva o negativa. Los otros dos métodos son descritos en la Guía del Instructor. Cuando este paso esté completo, la Herramienta MAGIC genera un archivo de disimilitud, el cual usted puede nombrar en la casilla "output file" (archivo de resultado). Al archivo se le dará automáticamente el sufijo ".dis". Haga clic en OK para empezar el proceso de computación. El progreso es monitoreado en una barra a escala que aparecerá en una ventana nueva (no se indica aquí). Usted puede calcular disimilitudes en cualquier archivo de expresión (.exp) pero usted debería utilizar sus proporciones transformadas en lugar de proporciones no transformadas. Usted también puede utilizar archivos de expresión normales y transformados que contienen valores de expresión absolutos. Debido a que la correlación y los cálculos a distancia no tienen significado alguno, si usted tiene menos de tres columnas, usted no podrá calcular disimilitudes si usted tiene dos o menos columnas.

# (13) Agrupe Genes

En este momento, usted puede generar una serie de agrupaciones utilizando cuatro métodos diferentes. El proceso de agrupación es uno muy popular para las micro-colecciones de ADN, así que describiremos éste primero, pero recuerde que la exploración es igualmente válida y le puede decir más acerca de sus genes y sus condiciones experimentales que en el proceso de agrupación. Explorar sus datos puede ser ejecutado cualquier tiempo después de la segmentación. Todo lo que usted necesita para explorar son los archivos de expresión (\*.exp).

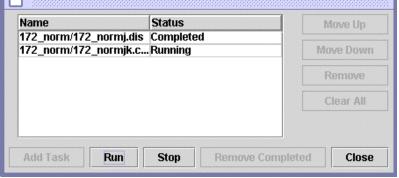
Con la Herramienta MAGIC, hay cuatro formas de agrupar genes. Usted puede agrupar desde cualquier archivo de disimilitud. Primero usted tiene que calcular las agrupaciones y después usted las puede mostrar en una variedad de formas. La forma más común de agrupación se llama la agrupación por jerarquía, la cual usted puede hacer con la herramienta MAGIC. Sin embargo, nosotros preferimos la agrupación Q-T (vea la Guía del Instructor para más detalles), aunque la Agrupación por Jerarquías es el único formato actualmente compatible con el programa de visualización de datos, Java TreeView. Usted también puede agrupar por "k-means" (método k) o "supervised clustering" (agrupación supervisada).

Una vez que usted ha agrupado los genes, usted puede mostrar los resultados de varias formas. La Herramienta MAGIC le permite ver estas agrupaciones en una variedad de pantallas dinámicas. Cada pantalla puede ser guardada como un archivo de imagen para publicación o fines didácticos. Las opciones para mostrar dichos resultados son discutidas en detalle más adelante en este manual.

#### Automatización de Tareas

Mientras más grandes sus grupos de datos se vuelven, el tiempo que le tomará hacer todos los cálculos necesarios incrementará rápidamente. Es por esto que la Herramienta MAGIC le permite establecer una lista de tareas que serán ejecutadas en secuencia. Usted le puede decir a la Herramienta MAGIC empezar una serie de pasos y después alejarse de su computadora. La Herramienta MAGIC ejecutará esta secuencia de tareas mientras usted hace otras cosas. Por ejemplo, usted puede establecer una lista de tareas para ejecutar e irse a casa para descansar. Cuando usted regrese la siguiente mañana, la Herramienta MAGIC habrá completado la serie de

tareas.



#### **Comentarios Finales**

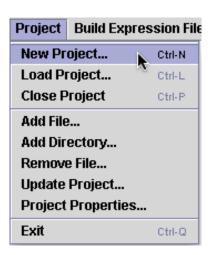
Esta sección ha sido diseñada como una plataforma para iniciarlo en la Herramienta MAGIC y su forma de trabajar con micro-colecciones de ADN. La Herramienta MAGIC le permite comparar las consecuencias de opciones diferentes para cuantificar, comparar y agrupar el mismo set de datos sin procesar. Esta capacidad de comparar métodos es una manera poderosa de comprender mejor las suposiciones y las implicaciones inherentes en el análisis de datos publicado cada semana. La Herramienta MAGIC le permite explorar datos y análisis de datos durante los primeros días de micro-colecciones de ADN cuando la comunidad científica todavía no establecido estándares para comparar resultados. La Herramienta MAGIC ha sido diseñada para darle poder al usuario y hacer las micro-colecciones de ADN más accesible para una audiencia más amplia y variada. En la siguiente sección. Cada opción disponible en la Herramienta MAGIC le será descrita para que usted pueda utilizar a su máximo la Herramienta MAGIC.

#### Lista Completa de Opciones de la Herramienta MAGIC

# Menú de Proyectos "Project Menu"

#### **New Project...** (Nuevo Proyecto...) (Control N)

Este comando empieza un nuevo proyecto. Un proyecto es una forma de organizar todos los trabajos relacionados a la Herramienta MAGIC en una carpeta. El nombre que usted le dé al proyecto es el nombre de la carpeta, y esta carpeta es automáticamente creado por la Herramienta MAGIC. Cada nombre de proyecto debe ser único y descriptivo. En la carpeta creado por la Herramienta MAGIC usted encontrará un archivo que termina con el sufijo ".grpj". Todos los pasos subsecuentes y los archivos serán guardados automáticamente en esta carpeta de proyecto, hasta que usted comience un proyecto nuevo. El archivo .grp es un archivo de texto que es esencialmente una tabla de contenidos (índice) de su proyecto.



# Load Project... (Cargar un Proyecto...) (Control L)

Esto le permite reabrir un proyecto previo. Navegue hacia el directorio donde se encuentra su proyecto en el disco duro, y seleccione el archivo .grpj en la carpeta del proyecto.

# Close Project (Cerrar el Proyecto) (Control P)

Este comando le permite parar un proyecto sin salirse de la Herramienta MAGIC completamente. Usted también puede parar un proyecto al abrir un proyecto nuevo y confirmar que usted quiere cerrar el proyecto que está actualmente abierto.

# Add File... (Agregar un Archivo...)

Este comando le permite agregar archivos de otros proyectos a su proyecto actual (por ejemplo archivos de extensión tiff, lista de genes, información de archivos, archivos de expresión existentes). Usted será dirigido a una ventana en la cual usted puede hacer clic, navegar por el disco duro buscando los archivos que usted quiere agregar. Si mantiene simultáneamente presionada la tecla Shift y hace clic en diferentes archivos, esto le permite seleccionar un rango consecutivo de archivos. (En Windows, usted puede mantener presionada la tecla Control y hacer clic en múltiples archivos).

# Add Directory... (Agregar un Directorio...)

Le permite agregar todos los archivos de la carpeta seleccionado a su carpeta actual.

# Remove File... (Eliminar un Archivo...)

Le permite eliminar archivos que no necesite en su carpeta de proyecto actual. Usted también puede borrar los archivos al escribir sobre la versión vieja (a usted se le preguntará si en verdad quiere sobrescribir sobre el archivo existente con el mismo nombre).

# **Update Project... (Actualizar un Proyecto...)**

Le permite arrastrar archivos a carpetas existentes y después actualizar el proyecto actualmente activo. Esto le ayuda al usuario a rápidamente mover archivos tiff, de cuadrículas, de expresión, de disimilitud y de agrupación alrededor y utilizarlos en proyectos diferentes.

# Project Properties... (Propiedades del Proyecto...)

Le permite modificar las propiedades predeterminadas y configurar la conducta de la Herramienta MAGIC. Hay tres lengüetas, cada una conteniendo propiedades de tipos diferentes.

**Data Handling (Manejo de Datos):** Actualmente la única opción de manejo de datos se relaciona a cómo manejar los datos ausentes. Usted puede escoger entre *eliminar* o *ignorar* cualquier gen en su proyecto actual que no tenga datos. Cuando una micro-

colección de ADN es impresa, algunas puntos estarán ausentes y por consecuente usted no puede recolectar datos para ese gen. Si usted decide *eliminar* todos los genes a los que les faltan datos, entonces los genes a los que les faltan cualquier dato de una o más columnas no serán utilizados para calcular disimilitudes. Si usted decide *ignorar*, a usted se le pedirá el porcentaje de datos posibles que deben estar disponibles para que un gen sea incluido en su análisis de datos. Esto le permite trabajar con genes a los cuales les faltan datos de menos de dicho porcentaje en las columnas. Los genes a los que les falten más de la cantidad mínima en porcentaje en las columnas, serán eliminados.

**Image Saving (Guardar Imagen):** Controla el tamaño máximo de la imagen guardado desde la Herramienta MAGIC.

Group Files (Agrupar Archivos): Hay dos opciones en esta lengüeta. La primera, "New Expression Files Carry Group Files When:" (Nuevos Archivos de Expresión Traen Consigo Archivos de Grupo Cuando) controla como archivos de grupo van mano a mano con los archivos de expresión. Esta opción se pone en juego cuando usted crea un nuevo archivo de expresión a partir de un archivo de expresión existente, por ejemplo al hacer una transformación de logaritmos, al añadir información, durante filtración, durante normalización o durante la filtración de datos. Ya que un archivo de grupo es simplemente una lista de genes, usted querrá asegurarse que los grupos que usted haya seleccionado en una versión anterior, selección basada en valores de los datos de expresión, sean accesibles después de que usted ejecute uno de los procesos mencionados anteriormente para crear un nuevo archivo de expresión. El ajuste predeterminado es "Always" (Siempre), lo que significa que todos los archivos de grupo son copiados a la carpeta que contiene el nuevo archivo de expresión. Usted también puede escoger "nunca escoger los archivos de grupo", o solamente copiar los archivos de grupo cuando los datos de expresión no han sido cambiados (por ejemplo cuando se añade información al archivo de expresión).

### Exit (Salir del Programa) (Control Q)

Este comando cierra la Herramienta MAGIC. Todos los pasos y archivos que se termino serán guardados a la carpeta de su proyecto. Los pasos incompletos serán eliminados. Los archivos tiff no serán reabiertos cuando el proyecto sea abierto después.

#### Menú "Build Expression File" (Construir Archivo de Expresión)

#### Load Image Pair... (Cargar Par de Imágenes...) (Control R and Control G)

Esto le permite explorar su disco duro para encontrar los archivos de extensión tiff para los dos colores. Usted puede cargar los dos archivos tiff en cualquier orden. Si



usted ha agregado archivos al proyecto, o si usted ha movido archivos hacia la carpeta de su proyecto y lo ha actualizado, todos los archivos tiff serán localizados en la carpeta "Images" (Imágenes) del proyecto. Si no es así, usted puede navegar hacia el directorio de los archivos en su disco duro. Tan sólo asegúrese de relacionar correctamente los colores con los archivos. Recuerde que el rojo es una longitud de onda más larga que la verde.

#### Load Gene List... (Control X)

Esta opción lee un archive que asocia cada punto en la micro-colección con un nombre de gen. La Herramienta MAGIC requiere que usted tenga este archivo, llamado la lista de genes, en un formato particular. La lista de genes, en el formato que usa la Herramienta MAGIC, son disponibles para ser descargadas de los sitios web del GCAT y de la Herramienta MAGIC, y vienen incluídas en los Sample Files (Archivos de Ejemplo), que también se pueden descargar desde el sitio web de la Herramienta MAGIC.

En varias ocasiones, listas de genes formateadas que no provienen de la Herramienta MAGIC tienen información adicional como por ejemplo qué puntos no han sido impresos, nombres alternativos para los genes, etcétera. Usted puede abrir su lista de genes para ver qué tipo de información contiene. Si dicha lista contiene información acerca de las placas y pozos para cada gen, ésta no es información útil para la Herramienta MAGIC pero ha sido utilizada para las personas que han ayudado a imprimir los chips puedan mantener un registro de lo que estaban haciendo durante la manufacturación de los chips.

Si usted tiene una lista de genes que no está en el formato de la Herramienta MAGIC, usted puede utilizar las instrucciones y ejemplos que se encuentran en este sitio web: <a href="http://www.bio.davidson.edu/peopple/macampbell/ACS\_MAGIC/genelists.html">http://www.bio.davidson.edu/peopple/macampbell/ACS\_MAGIC/genelists.html</a> para crear una lista de genes con el formato apropiado. Primero, usted deberá abrir su lista de genes desde Excel. Encuentre la columna que contiene nombres de tipo ORF como YBL023c o YAR002W, etcétera. Copie esta columna ORF y péguela en la primera columna (usted posiblemente tendrá que crear una columna nueva para mantener esta información). A continuación, elimine todas las filas representan el título, de forma que la primera fila en su archivo sea el primer gen en la lista. Guarde el archivo modificado como un archivo de texto delimitado por la lengüeta, con un nombre nuevo que termine con el sufijo ".txt" (archivo de texto). Este archivo es ahora una lista de gen válida para ser utilizado en la Herramienta MAGIC. Aunque toma un poco de trabajo manual crear dicha lista de genes compatible con la Herramienta MAGIC, esto le permite al usuario adaptarse rápidamente a diferentes estilos de producción de micro-colección Más adelante, usted aprenderá cómo importar información de genes de organismos comúnmente estudiados.

# "Load Saved Grid" (Cargar Cuadrícula Guardada) (Control W)

Esta opción en el menú le permite que usted abra, en la Herramienta MAGIC, una cuadrícula que usted previamente la había guardado, la misma que fue creada cuando escojió la opción que se indica a continuación "Create/Edit Grid" (Crear/Editar Cuadrícula).

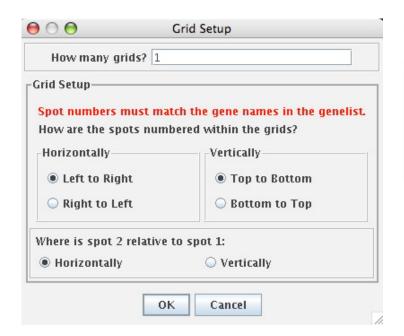
# "Combine and Load Grid Files" (Combine y carge archivos de cuadrícula)

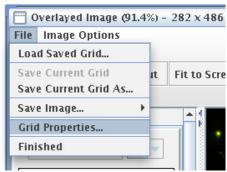
Si usted crea dos archivos de cuadrícula diferentes, usted puede combinarlos utilizando esta opción. Cuando usted escoja esta opción del menú, se le pedirá que escoja el primer archivo de cuadrícula. Desde este archivo, la Herramienta MAGIC tomará los detalles de la orientación de la cuadrícula que usted determinó en el paso (b) mencionado anteriormente, además de tomar todas las cuadrículas en este archivo. Una vez que usted seleccione el primer archivo de cuadrícula, se le pedirá entonces seleccionar el segundo archivo de cuadrícula y entonces el nuevo nombre de archivo para el archivo combinado de cuadrícula. La Herramienta MAGIC toma las cuadrículas del primer archivo como las primeras cuadrículas n en el nuevo archivo, seguido de las cuadrículas del segundo archivo como el resto de las cuadrículas. Usted deberá asegurarse que las cuadrículas estén combinadas en el orden correcto. Una vez que el archivo de cuadrícula haya sido creado, la Herramienta MAGIC cargará de forma automática el archivo de cuadrícula combinado, y después usted podrá editar la cuadrícula al escoger la opción "Create/Edit Grid" (Crear/Editar Cuadrícula), o continuar directamente hacia la Segmentación.

#### Create/Edit Grid (Crear/Editar Cuadrícula) (Control A)

Cuando usted empiece el proceso de "addressing" (distribución) y "gridding" (cuadricular), usted deberá ver primero una imagen combinada de sus archivos de expresión rojos y verdes, y donde el rojo y el verde estén superimpuestos, usted deberá observar una sombra amarilla. Es entonces cuando se le formularán cuatro preguntas para indicarle a la Herramienta MAGIC cómo están los puntos enumerados, ésto está indicado en la ilustración abajo. Este paso, llamado "addressing" (distribución), es donde usted podría equivocarse fácilmente, así que por favor tenga cuidado cuando usted conteste las cuatro preguntas que aparecen en la ventana. Es vital que usted entienda la forma en la que sus puntos están organizados en su micro-colección y en la lista de genes. Todas las preguntas deberán ser contestadas según la forma en la que usted vea la imagen combinada de su micro-colección en la ventana. ¿Los genes han sido impresos doble? Si ése es el caso, ¿los puntos duplicados son horizontales o verticales? Usted deberá saber cuántas cuadrículas existen así como el orden de los puntos en su lista de genes comparados con la imagen en la Herramienta MAGIC. Las respuestas predeterminadas en las preguntas de "Grid Setup" (Ajustes de la Cuadrícula) corresponden a la forma en la que usted leería un libro: de izquierda a derecha, de arriba a abajo, con el segundo punto estando localizado de forma horizontal del primero. No se puede hacer suficiente énfasis cuan crítico este paso es. Si usted se equivoca en esta parte, usted no tendrá la correcta identidad de sus puntos. Una vez que usted

presione OK, usted habrá terminado el paso "addressing" (distribución), pero usted puede siempre ir a "File, Grid Properties" (Archivo, Propiedades de la Cuadrícula) en la ventana Gridding para contestar de nuevo estas cuatro preguntas.





"Gridding" (Cuadricular) es mucho más fácil. El propósito de cuadricular es aquél de dibujar casillas pequeñas alrededor de cada punto de forma que éstos estén en el centro de las casillas. Usted encontrará útil hacer un acercamiento en la primera cuadrícula de puntos. Para hacer dicho acercamiento, haga clic en el botón "Zoom In" (Acercar) y después haga clic donde sea que usted quiera se haga un acercamiento al centro. La lengüeta con el número uno deberá ser seleccionada como la predeterminada.

Navegue por la imagen hasta que usted pueda ver la primera cuadrícula que usted sabe es la primera en el diseño original de su micro-colección. Si usted así lo desea, puede ajustar el contraste para ayudar esclarecer un poco los puntos tenues. Para hacer esto, deslice el indicador que está actualmente apuntando hacia 100% de contraste cerca de la parte superior de la ventana. Si el valor máximo del deslizador todavía no le da suficiente contraste, usted puede hacer más ajustes al escribir en la casilla junto al deslizador el porcentaje de contraste que usted quiere. Ajustar el contraste NO afecta los datos sin procesar, tan sólo le permite ver los puntos de mejor forma para este paso.

Para cuadricular, tan sólo haga clic en tres puntos. Primero, haga clic en el botón que dice "Set Top Left Spot" y después haga clic en el centro del punto de la izquierda superior en la cuadrícula. Segundo, haga clic en el botón que dice "Set Top Right Spot" y después haga clic en el centro del punto de la derecha superior en la cuadrícula. Tercero, haga clic en el botón que

dice "Set Bottom Row" y después haga clic en el centro de cualquier punto en la fila inferior. Elija un punto redondo grande para hacer este paso más fácil. Escriba el número de filas y columnas. Esto será contestado basándose en la forma en la que usted está actualmente viendo el archivo tiff. En este ejemplo, hay 24 filas y 12 columnas. Haga clic en el botón "Update" (Actualizar). En este momento, usted deberá poder ver todos los puntos en la primera cuadrícula rodeados por las casillas como se indica a la derecha. (Es posible que usted deba hacer un "zoom out" (alejamiento) para ver la cuadrícula completa).

En este momento, vea si los puntos se ven centrados en las casillas. Si éste no es el caso, entonces ajuste la posición de las casillas ya sea haciendo clic en el botón apropiado y después en el punto correcto, escribiendo los números correctos para ajustar las casillas, o ajustando la cuadrícula con el ratón. Si usted hace clic en cualquier parte del interior de la cuadrícula, usted puede arrastrar toda la cuadrícula a una localización nueva. La cuadrícula puede ser reajustada en tamaño desde una esquina al hacer clic en uno de los puntos grises y arrastrando el ratón. Mientras usted hace esto, el nuevo tamaño y la posición de la cuadrícula serán indicados. Finalmente, si usted hace clic en uno de los botones de rotación, toda la cuadrícula va a rotar alrededor de su eje central, permitiéndole ajustar cuadrículas que esten un tanto inclinadas en sus imágenes. Si usted decide manualmente ajustar la cuadrícula al cambiar los valores en las casillas, tenga en cuenta que la posición del ratón se indica en la esquina izquierda de la ventana de forma que usted puede determinar si los números deberían ser más grandes o más pequeños para ajustar las casillas en la dirección correcta. Hacer el proceso de cuadricular (Gridding) toma un poco de práctica, pero es MUCHO más fácil que la mayoría de los métodos de Gridding.

Una vez que la primera cuadrícula ha sido propiamente rodeada con las casillas y con los puntos en cada respectivo centro, es tiempo de repetir este proceso para la cuadrícula #2. Seleccione en el teclado la tecla control (Ctrl) y manténgala apretada

Mientras selecciona en la parte superior izquierda de la cuadrícula #2, el centro del punto. La misma cuadrícula, trasladada hacia la localización especificada por el clic de su ratón, aparecerá como la cuadrícula #2, y todos los números en las casillas en la izquierda serán llenados automáticamente. Si usted deja de presionar la tecla Control, usted puede ajustar la cuadrícula como usted lo hizo previamente. Repita este proceso para todas las cuadrículas. Cada vez que usted haga clic mientras mantiene presionada la tecla Control, usted colocará automáticamente el siguiente número de menor valor que todavía no haya sido definido como cuadrícula. Continúe este proceso hasta que todas las cuadrículas estén rodeadas de casillas.

Si usted necesita mover múltiples cuadrículas a la vez, presione y mantenga la tecla Shift, y después haga clic en cada cuadrícula que usted desee mover. Al seleccionar las cuadrículas, éstas serán resaltadas en color azul. Una vez que todas las cuadrículas que usted quiera mover estén visualizadas en azul, haga clic y arrastre adentro de cualquiera de las cuadrículas para moverlas

todas a la vez. Usted también puede rotar cuadrículas múltiples a la vez al seleccionarlas de la misma forma y hacer clic en uno de los botones de rotación.

Usted puede guardar su cuadrícula actual en cualquier momento, utilizando "File, Save Current Grid" (Archivo, Guardar Cuadrícula Actual) o "Save Current Grid As..." (Guardar Cuadrícula Actual Como...) para guardarla con un nombre diferente. Los archivos de cuadrícula son asignados automáticamente el sufijo ".grid". Usted también puede cerrar la ventana de cuadriculación sin guardar sus cambios, y la cuadrícula actual será restablecida automáticamente la próxima vez que usted abra la ventana de cuadriculación (sin que se le formulen las cuatro preguntas otra vez). Sin embargo, si usted cierra el proyecto, usted deberá guardar su cuadrícula antes de cerrar su proyecto, y escoger la opción "Load Saved Grid" (Cargar Cuadrícula Guardada) cuando usted empiece a cuadricular la próxima vez. Esto le permite retomar el trabajo que usted empezó anteriormente en el proceso de cuadriculación.

Cuando usted haya acabado el proceso de "Gridding" (cuadricular) en todas sus matrículas, haga clic en el botón "Done!" (¡Listo!). Si usted no ha guardado todavía su cuadrícula, el programa le pedirá hacerlo antes de poder avanzar al siguiente paso. Si el número de genes en su lista de genes y el número de puntos que usted ha cuadriculado no coinciden, aparecerá un mensaje de error. Usted deberá tener exactamente una casilla de la cuadrícula por cada línea (ya sea un gen o réplica de gen) en la lista de genes. De no ser así, usted cometerá probablemente un error al momento de identificar los puntos más adelante, así que se le pide arreglar el problema ahora. Si su lista de genes y el número de puntos coinciden, se le informará el número total de puntos y se le permitirá guardar el archivo de la cuadrícula para uso posterior.

Usted también puede guardar una foto de la pantalla de las imágenes combinadas de extensión tiff en cualquier momento antes, después o durante el proceso de cuadricular (Gridding). Usted puede guardar la imagen de lo que quiera que actualmente se indique en el interior de la ventana de cuadricular con un formato tiff, jpg o gif. (El formato tiff funciona en todos los procesadores de texto y de dibujo lo que lo hace un formato universal. El formato jpeg es una buena opción para imágenes como éstas que pueden contener muchas sombras, de la misma forma que uno utilizaría el formato jpeg para una fotografía. El formato gif es el más simple lo cual puede hacer que usted pierda la sutileza de los detalles en el archivo original). Guardar esta imagen combinada es útil si usted quiere tomar una fotografía de toda la cuadrícula, la cual puede utilizar para publicaciones o fines didácticos.

Spot Flagging (Determinar los puntos problemáticos) (Control F)

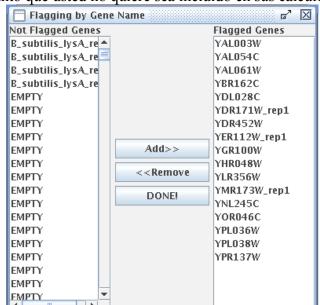
Si hay puntos en su cuadrícula que usted no desea que sean utilizados en su análisis de datos, usted puede escoger excluir los datos en este momento, antes de la creación del archivo de

expresión. Para hacer esto, elija "Spot Flagging" del submenú "Addressing/Gridding", o presione Control + F.

De la misma forma que en la ventana gridding, usted puede hacer un zoom in (acercamiento) o un zoom out (alejamiento), y ajustar la imagen a la pantalla. También como en la ventana de la cuadrícula, cuando usted desplaza el ratón sobre un punto, la barra de estado en la parte inferior de la ventana le indicará la información acerca del gen. Si usted ve un



punto que usted no quiere sea incluido en sus cálculos, haga clic sobre dicho punto. Una cruz



"X" de color azul aparecerá sobre el punto marcándolo como "flagged" y el proceso de segmentación lo ignorará.

Si usted ha determinado opciones de marcación automáticas y ha calculado datos para los puntos, cruces de color naranja aparecerán encima de los puntos automáticamente marcados. Estas marcaciones automáticas sólo pueden ser alteradas al cambiar las opciones de marcación automáticas en la ventana de Segmentación.

Para ver cuáles genes han sido "flagged", o para elegir los genes que serán marcados o

aquéllos que no deberán serlo en función de su nombre, seleccione "Flagging by Gene Name" del menú Flagging. En el diálogo que aparece, los genes no marcados (aquéllos que serán usados) aparecen en la izquierda de la pantalla, y aquéllos que son marcados aparecen en la derecha. Para marcar un gen, haga clic (y por ende resaltándolo en la lista) de la izquierda y haga clic en "Add" (Agregar). Para desmarcar un gen, haga clic en su nombre en la lista de la derecha y después haga clic en "Remove" (Quitar). Usted puede seleccionar múltiples artículos en la lista

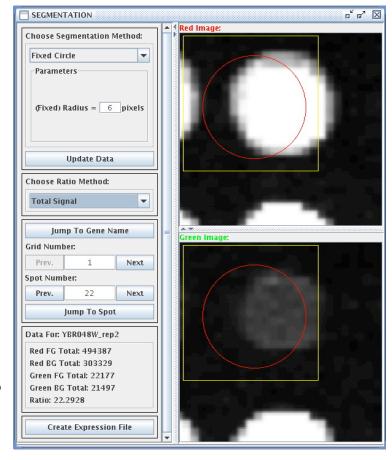
al mantener presionada la tecla "Control", y después al hacer clic en cada artículo, o, para seleccionar un rango de sujetos, haga clic en el primero, después mantenga presionada la tecla "Shift", y después haga clic en el último artículo. Una vez que usted haya presionado el botón "Add" o "Remove" (Agregar o Quitar), los cambios se vuelven visibles en la imagen detrás de la ventana "Flagging by Gene Name" (marcador por nombre de gen). Note que los genes con nombres "empty", "missing", "none", o "blank" (en español: vacío, ausente, ninguno, o blanco respectivamente) son automáticamente excluidos del archivo de expresión, así que no necesitan ser marcados por nombre. Cuando usted haya acabado de marcar genes por nombre del gen, haga clic en "Done!" (¡Listo!).

De la ventana principal "Flagging" (Marcación), usted también puede ya sea cargar o guardar archivos de marcación. Estos archivos tiene la extensión ".flag" y se encuentran guardados en la subcarpeta titulada "flags" en la carpeta principal que contiene su proyecto. Este proceso de guardar su trabajo trabaja de la misma forma que aquél descrito para el cuadriculado, excepto que en este proceso a usted no se le pide guardar un archivo de marcación automáticamente. Para cargar un archivo de marcación, abra la ventana "Spot Flagging" (Marcación de Puntos) y después elija "Load Saved Flags" (Cargar Marcaciones Guardadas...) del menú File (Archivo). En esa ventana, usted puede escoger el archivo de marcación que usted quiera cargar. Note que el número de cuadrículas y el número de puntos por cuadrícula tiene que ser el mismo de la cuadrícula actual para que usted pueda cargar el archivo de marcación.

# **Segmentación (Control S)**

Segmentación es el proceso de distinción de señales del fondo. Hay tres métodos disponibles para este proceso. Durante la segmentación, usted tendrá la oportunidad de ver cada punto en toda la micro-colección, una a la vez. En este paso, los dos archivos tiff son separados otra vez, con la imagen roja arriba y la imagen verde abajo. Hay tres algoritmos disponibles en la Herramienta MAGIC para encontrar, por separado, el primer plano (señal) y el fondo (ruido) en cada canal (rojo y verde). Además, hay cuatro opciones para combinar estos cuatro valores numéricos para determinar la proporción.

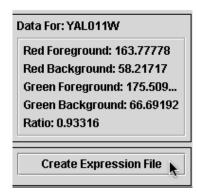
Se le recomienda experimentar con los algoritmos y opciones diferentes antes de escoger el mejor método. Al explorar de punto a punto, o saltar hacia un punto potencialmente problemático que usted identificó mientras usted hacía el proceso de cuadriculación, usted puede ver cómo estas opciones afectarán su respuesta final. Cuando usted esté satisfecho con sus opciones, haga clic en el botón "Create Expression File" (Crear Archivo de Expresión), y a usted se le pedirá escoger un archivo en el cual la Herramienta MAGIC guardará todas las proporciones una para cada punto en la micro-colección. Cuando usted guarde toda la lista, todos los valores serán computados



nuevamente, así que no importa si usted ha explorada dos puntos o doscientos. Además de guardar la lista de proporciones, a usted se le dará la oportunidad de guardar los datos "sin procesar" ("raw data"), es decir, todos los valores tanto en el primer plano como en el fondo en los canales verdes y rojos.

#### Círculo fijo

El círculo fijo simplemente posiciona un círculo en medio de las casillas. Todos los pixeles dentro del círculo (que también están dentro de la caja) serán considerados como señal y los pixeles que están afuera del círculo (pero que están todavía dentro de la caja) serán considerados como fondo. Usted puede establecer el radio del círculo en unidades en pixeles. En la figura de arriba, usted puede ver los puntos que están en la casilla, pero éstas no están centradas. Los valores de primer plano y del fondo que están descentralizados y los puntos que son más grandes o más



chicos que el radio seleccionado fijo no serán exactamente correctos. Sin embargo, la proporción entre los valores rojos y verdes deberán ser bastante precisos. El círculo fijo es el método más

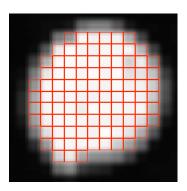
común para hacer la segmentación, y es el método más rápido de los tres métodos de segmentación.

# Círculo adaptivo

Este método cambia el centro y el radio del círculo para entallar el tamaño y la localización de cada punto. El algoritmo considera todos los pixeles de valor más grande que aquél establecido por el usuario como "on" (prendidos), y encuentra el círculo con el porcentaje más grande de pixeles que están "on" (prendidos). El radio puede tener un rango entre un límite menor y mayor establecido por el usuario el centro puede estar donde quiera dentro de la casilla de la cuadrícula. Este método es más lento que el método del círculo fijo, pero en general cubre el punto de mejor forma.

#### Crecimiento de Región Asemillada

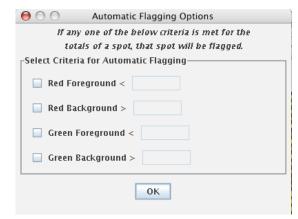
Este método de segmentación está diseñado para encontrar la señal en cada punto basado en la distribución de la señal. Este método de segmentación busca el pixel más brillante y después conecta todos los pixeles adyacentes a dicho pixel y los conecta en una forma. El algoritmo conecta de forma simultánea los pixeles al fondo y a la región de primer plano, continuando hasta que todos los pixeles estén en una de las regiones. Un límite especificado por el usuario determina cuáles pixeles pueden ser utilizados para "asemillar" la región. Éste es el método más lento ya que cada pixel es procesado individualmente.



Sin importar qué método usted escoja, usted puede inspeccionar visualmente los puntos para verificar que el cuadriculado y que la segmentación hayan sido ejecutadas adecuadamente. Esta inspección le da la oportunidad de marcar cualquier punto que usted crea no debe ser

considerado durante análisis de datos posteriores.

Una vez que usted haya seleccionado su método de segmentación y de proporción, usted puede establecer criterios de forma que si cualquier punto no cumple con dichos requerimientos, su proporción no será incluida en el archivo de expresión. Para hacer esto, haga clic en el botón "Automatic Flagging Options" (Opciones Automáticas de Marcado). Aquí, usted puede determinar valores límites para estos criterios, y



escoger marcar un punto ya sea si uno (operador Boolean OR) o todos (operador Boolean ALL) los requisitos han sido cumplidos para dicho punto. Cuando usted haga clic en OK (incluso en el

caso de que se deje en blanco los valores límites), se le preguntará si usted desea ejecutar cálculos o no para encontrar el estado de marcación de los puntos. En el proceso, la Herramienta MAGIC también computa el promedio y la deviación estándar para cada uno de los cuatro puntos de información utilizados en los cálculos (aun cuando usted ha dejado los valores límites en blanco). Usted puede entonces utilizar estos datos para refinar los requisitos de marcado automático. Por ejemplo, usted podría desear marcar aquellos genes cuyo primer plano total rojo o verde sea menos de dos deviaciones estándar debajo el promedio.

Red FG Average: 2005775.9436
Red FG Std. Dev.: 1556768.4052
Red BG Average: 1343939.1146
Red BG Std. Dev.: 1047404.7682
Green FG Average: 1640906.6979
Green FG Std. Dev.: 1469862.3653
Green BG Average: 1046154.8307
Green BG Std. Dev.: 948306.8859

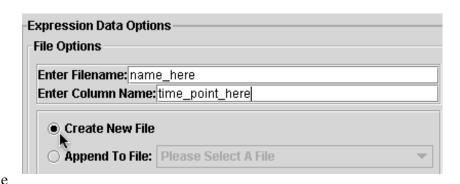
Para ver en una cuadrícula cuáles puntos han sido marcados, abra la ventana "Spot Flagging" del submenú "Addressing/Gridding". Todos los puntos que han sido automáticamente marcados serán marcados con una "X" naranja. Estas banderas pueden ser únicamente cambiadas al ajustar los requisitos de marcado automáticos, pero usted puede agregar o eliminar marcaciones automáticas en este momento también. Si

un punto es marcado de forma automática y manual, una "X" azul aparecerá superimpuesta en el punto en lugar de la "X" naranja. Si usted desmarca de forma manual un punto que ha sido previamente marcado automáticamente, la "X" se volverá de color naranja y el punto seguirá siendo contado como marcado. Si usted ajusta las opciones de marcado automáticas, usted deberá calcular de nuevo los datos para hacer que las marcaciones automáticas revisadas aparezcan en la pantalla "Spot Flagging" (Marcación de Puntos).

Usted también puede crear diagramas MA y RI "ratio-intensity" (proporción-intensidad). Estos diagramas le ayudarán a visualizar qué tan uniforme la impresión y la hibridación en su chip ha sido, y también le puede ayudar a determinar si usted necesita ejecutar un tipo de normalización fuera de la Herramienta MAGIC.

Nota: en este contexto,  $M = \log_2 R/G$ ,  $A = \log_2 RG$  insert square root character here

Cuando usted complete la segmentación, usted producirá un archivo de expresión. Haga clic en "Crear Archivo de Expresión", cuando usted esté satisfecho con el proceso de segmentación. Esto generará un archivo de



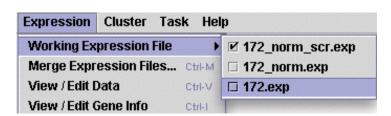
expresión, el cual ha sido el objetivo de la primera mitad de la Herramienta MAGIC. Un archivo de expresión contiene las proporciones para cada punto (rojo dividido entre verde). Una

proporción de 999 significa que una división entre cero hubiera ocurrido, lo que significa que la intensidad verde fue cero o negativa; una proporción de 998 significa que un 0 entre 0 hubiera ocurrido, lo que significa que ambas intensidades (roja y verde) fueron cero o menores que 0. La Herramienta MAGIC ignorará ciertos datos en la columna "gene name" (nombre de gen) ("blank", "EMPTY", "missing" and "none" (en español estos términos significan "en blanco" "VACIO" "ausente" y "ninguno") ya sea que estén escritos en mayúsculas o en minúsculas), y omitirá completamente cualquier punto marcado del archivo de expresión. Esto significa que para poder combinar los archivos correctamente, usted necesitaría marcar los mismos genes en todos los archivos de expresión que usted quiera combinar. Las proporciones serán utilizadas para todos los análisis de datos posteriores. Usted ya no necesita los archivos de expresión tiff.

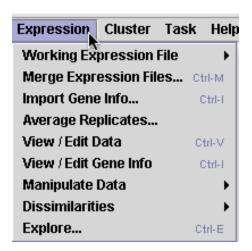
Usted necesitará nombrar el archivo de expresión y la columna (por ejemplo período de tiempo, tratamiento, etcétera). Usted puede añadir este archivo a un archivo previo o crear un nuevo archivo. Usted también puede guardar la señal "cruda" (datos sin procesar) y los niveles de intensidad en el fondo.

# Menú de Expresión

#### "Working Expression File" (Archivo de Expresión Activo)

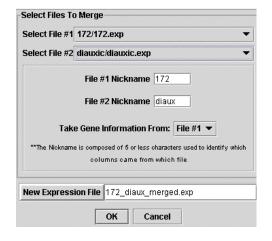


Esta opción le permite escoger entre un rango de archivos de expresión con los cuales usted puede trabajar dentro de un proyecto. Como usted puede apreciar en la imagen en la izquierda, usted puede escoger el archivo de expresión activo con sólo hacer clic en él.



# "Merge Expression Files..." (Unir Archivos de Expresión) (Control M)

Al unir los archivos de expresión, esta opción le permite combinar datos de múltiples chips para que usted puede evaluar datos de tipo "time course" u otros sets de datos relacionados. Usted une archivos de uno en uno y les asigna nombres para ayudar a la Herramienta MAGIC a llevar la cuenta de los datos que serán combinados próximamente. De la misma



forma, usted puede seleccionar un set de anotaciones de genes como la referencia que será mantenida en el set de datos combinados. Un nuevo archivo será creado, de forma que sus archivos originales no se perderán.

# "Import Gene Info..." (Importar Información de Gen) (Control R)

Esto le permite recolectar información más completa acerca de sus ORFs. Por ejemplo, hemos creado un archivo de texto que describe la localización en los cromosomas, las tres categorías de la ontología de anotación de genes, y sinónimos de todos los genes de levadura. Esto le permite buscar en cada uno de estas áreas para ayudar detectar tendencias e información relevante.

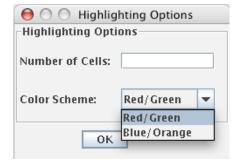
# "Average Replicates..." (Replicados Promedio)

La Herramienta MAGIC trata cada punto como único y no computa el promedio para genes replicados automáticamente. Esto preserva todos los datos originales de proporción. Si un set de nombres de puntos es idéntico en la lista de genes, la Herramienta MAGIC le dará a cada replicado un nombre único al adjuntar "...rep1" "...rep2", etcétera. Después de haber creado archivos de expresión, usted podrá escoger computar un promedio de los puntos replicados definidos por nombre ORF. Cuando usted promedie los **puntos** replicados, todos los puntos con nombres idénticos (sin tomar en cuenta el...rep#) serán entonces los datos promediados.

## "View/Edit Data" (Ver/Visualizar Datos) (Control V)

Después que un archivo de expresión es creado o combinado, usted puede ver y editar los datos. Esta opción no debería ser utilizada a menudo, pero nosotros hemos querido que usted tenga acceso a los datos de proporción si usted considera necesario hacer esto. Es útil si usted quiere verificar los pasos o retomar un proyecto después de un período prolongado de tiempo. Desde esta tabla, usted puede escoger resaltar las proporciones n más grandes o más pequeñas en cada columna de su archivo de expresión. Para hacer esto, escoja "Highligh Top and Bottom Ratios" (Resaltar Las Proporciones de las Partes más grandes o más pequeñas) del menú Edit (Editar). En la ventana "Opciones" que aparece, entre el número de células de proporción altas/bajas que

usted quiere que sean resaltadas, escoja el esquema de color, y haga clic en OK. Por ejemplo, si usted escribe "10" en la casilla y escoge rojo/verde como su esquema de color, las 10 células de valor más alto en cada columna serán resaltadas en rojo, y las 10 células más bajas en cada columna serán resaltadas en verde. Esta opción es útil para revisar la capacidad de reproducción entre experimentos.



# View/Edit Gene Info (Ver/Editar Información de Genes) (Control I)

Esta opción le permite ver y modificar las anotaciones de genes. Por supuesto, usted puede ver y editar este archivo fuera de la Herramienta MAGIC, pero esta opción le da una oportunidad de hacerlo en MAGIC. Usted también podría con esta opción llevar a cabo una búsqueda relacionada a la función de un gen. Mirando la lista le permitirá escoger términos apropiados para determinar la búsqueda.

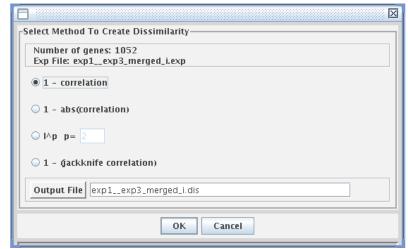
#### Replace Names With Aliases (Remplazar Nombres con Alias)

Si usted ha importado la información de genes hacia su archivo de expresión archivo, es probable que contenga diferentes alias, o nombres comunes, para los genes. Usted puede utilizar estos alias al escoger "View/Edit Gene Info". Por ejemplo, el alias del gen YBR167C es POP7. Esta opción también le permite remplazar los nombres de genes definidos en la lista de genes con su alias en el archivo de información, creando un nuevo archivo de expresión en el proceso. El antiguo nombre de gen será el alias en el nuevo archivo de expresión. Si el alias aparece más de una vez, a cada ocurrencia se le añadirá el sufijo "...repX" donde X representa un número comenzando del 1 al número de veces que dicho alias ocurra. Si un gen no tiene un alias, el nombre no cambiará.

#### "Dissimilarities" (Disimilitudes) (Control D)

Calcular las disimilitudes le permite comparar diferentes genes entre ellos. Cuando usted hace esto, una ventana aparecerá donde usted tendrá que escoger entre tres opciones. El método más común es la opción predeterminada, la correlación -1 (por favor vea la Guía del Instructor para

una explicación detallada y otros métodos). Cuando este paso esté completo, la Herramienta MAGIC generará un archivo de disimilitud el cual usted podrá nombrar en la casilla "Output File", cuya extensión será automáticamente ".dis". Haga clic en OK para empezar este proceso. El proceso es monitoreado en una barra de progreso que aparecerá



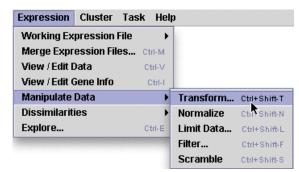
(no se indica aquí). Ya que la correlación y los cálculos a distancia no tienen sentido a menos que hayan al menos tres columnas, a usted no se le permitirá calcular disimilitudes si usted tiene dos o menos columnas.

### "Manipulate Data" (Manipular Datos)

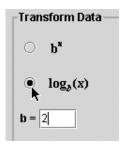
Manipular datos no es tan malo como se escucha. Esta opción le permite escoger entre cinco opciones. Estas opciones NO alteran sus datos originales, simplemente le permiten procesar más sus datos antes de hacer agrupaciones o explorar.

# Transform (Transformar) (Control Shift T)

Un proceso estándar que usted deberá ejecutar es transformar sus datos antes de llevar a cabo cualquier análisis (explorar o calcular disimilitudes y agrupar). Usted querrá hacer una transformación de tipo log a sus proporciones para eliminar cualquier fracción. Es importante obtener todas las proporciones en la misma escala de magnitud. Por ejemplo si un gen ha



sido reprimido 16 veces, la proporción será .0625 mientras un ge inducido 16 veces tendrá una proporción de 16.0. Antes de analizar sus datos, usted debe hacer una transformación logarítmica de sus datos. Después de la transformación (típicamente log2), los dos genes serían alterados (-4 vs. +4) con magnitud igual pero en direcciones opuestas. Vea la Guía del Instructor para más información. Usted deberá explorar antes de transformar, pero podría querer o no querer normalizar antes de explorar (vea abajo). Si usted quería "des-transformar" sus datos transformados, usted puede utilizar la función exponencial b<sup>x</sup>.



#### Normalize (Normalizar) (Control Shift N)

Este proceso toma sus proporciones transformadas y los corrige por la magnitud de las proporciones de un gen y la variación entre las proporciones de cada gen. La normalización no es apropiada por los datos de proporciones, pero es útil para valores de expresión absolutos. Vea la Guía del Instructor para más detalles.

#### Reorder/Delete Columns (Reordenar/Eliminar Columnas) (Control Shift L)

Si usted ha combinado datos de muchas micro-colecciones (por ejemplo, un experimento de tipo "time course"), usted podría querer estudiar solamente ciertas porciones de sus datos combinados de forma independiente. Limitar los datos le permite seleccionar los encabezados de las columnas y mantener estos datos seleccionados para un análisis en un set de datos limitados. Su archivo original de combinación permanecerá sin cambios y un archivo nuevo será creado. El

nuevo archivo de expresión tendrá una terminación "x\_limited.exp" donde x representa el nombre del archivo de expresión original.

*Filter (Filtrar) (Control Shift F)* 

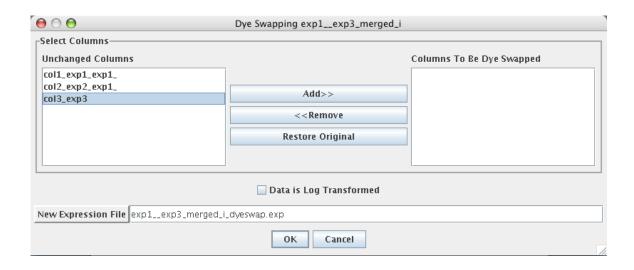
El filtrado de datos le permite eliminar consideraciones futuras de los genes que cumplen o no cumplen ciertos criterios definidos por el usuario. El filtrado se puede ejecutar a través de este menú o se puede llevar a cabo al guardar los resultados de cada búsqueda como archivos de expresión desde la ventana del Explorador (vea abajo).

Scramble (Revolver)

Le da usted tres métodos diferentes para crear un archivo de expresión de genes con los números exactos que su archivo actual, pero en orden aleatorio. La aleación puede ayudarle a indicar si los patrones encontrados a través de la exploración y la agrupación son efectos reales de las condiciones experimentales o no.

Dye Swap Data Manipulation (Manipulación de Datos A Través del Intercambio de Tinta) (Control Shift D)

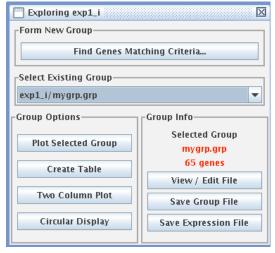
Si usted ha intercambiado las imágenes rojas y verdes mientras usted construía su archivo de expresión, usted puede intercambiar las proporciones después de la segmentación al escoger "Dye Swap Data Manipulation" del menú "Expression" (Expresión). Desde esta ventana, usted puede escoger columnas del archivo de expresión actual para que éstas sean intercambiadas de color de tinta. Si la casilla "Data is Log Transformed" (Datos son Tranformados por Logaritmos) no está activada, entonces las proporciones de las columnas seleccionadas serán recíprocas para obtener los nuevos valores. Si esta casilla está activada, los datos serán negados para obtener los nuevos valores (ver figura en la página siguiente).



# "Explore" (Explorar) (Control E)

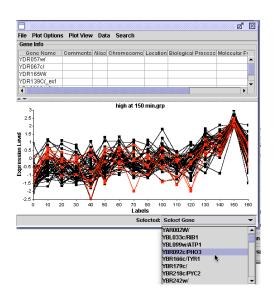
Después de haber transformado sus datos, usted puede explorarlos de diferentes maneras. El

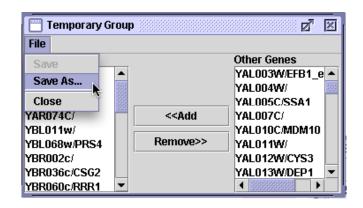
grupo predeterminado de genes es la lista completa en el archivo de expresión. Usted puede seleccionar un subset de genes a través del botón "Form New Group" (Formar Nuevo Grupo) llamado "Find Genes Matching Criteria..." (Encontrar Genes Que Cumplan con los Criterios...) Usted puede buscar requerimientos similares a aquéllos mostrados para el set de filtración en la página previa. Cuando usted haya identificado genes de su interés, la ventana cambia como se muestra en la derecha en el texto en rojo. Para guardar este nuevo grupo de genes, haga clic en el botón "View/Edit File" (Ver/Editar



Archivo) que se encuentra justo abajo el texto en rojo, o haga clic en el botón "Save Group File" (Guardar Archivo de Grupo) que se encuentra justo abajo del botón previamente descrito. Después de guardar un nuevo archivo de expresión, se le preguntará si usted quiere explorar el nuevo archivo o mantener el antiguo archivo abierto. Si usted abre el nuevo archivo, usted puede utilizarlo para una construcción continua de búsquedas – en el nuevo archivo de expresión, formar un nuevo grupo al hacer clic en el botón "Find Genes Matching Criteria..." y usted puede hacer la búsqueda en el nuevo archivo de expresión.

Una nueva ventana aparecerá que le permite ver la lista de genes en su grupo recién formado. Usted puede modificar este grupo si así lo desea, o usted lo puede guardar haciendo clic en el comando "save as" en el menú File (Archivo). Usted puede crear varios subgrupos de genes y explorarlos individualmente utilizando el menú "Select Existing Group" (Seleccionar Grupo Existente). Una vez que usted tenga subsets de genes para explorar, usted los podrá visualizar de diferentes maneras:



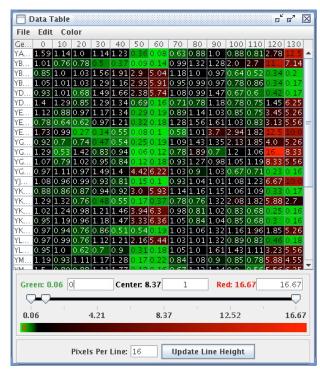


Plot Selected Group (Graficar Grupo Seleccionado)

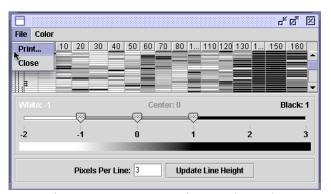
Usted puede obtener un diagrama de las proporciones mostradas en forma gráfica. Usted puede seleccionar un gen utilizando el menú en la esquina inferior derecha. O, como se muestra aquí, usted puede hacer clic en un nódulo a la vez y mantener presionada la tecla Shift para seleccionar genes múltiples (en este caso, aquéllos con las proporciones de menor valor en el grupo). Estos genes seleccionados son enlistados en la ventana superior (la cual usted puede arrastrarla hacia abajo para ver) así como cualquier otra información acerca de estos genes en su lista. Usted puede ajustar el tamaño del diagrama, así como hacer un acercamiento en alguna sección. Por ejemplo, este grupo de genes ha sido seleccionado por tener una proporción de 2 o más a un tiempo de 150 minutos. Para desenredar las líneas amontonadas, usted puede hacer un acercamiento sobre cualquier región de interés. Para hacer esto, mantenga presionada la tecla Control y después haga clic y arrastre una casilla alrededor del área amontonada para hacer un acercamiento. Usted puede hacer un alejamiento utilizando el menú Plot View en la parte superior de la ventana. Además, usted puede nombrar los axis, guardar este diagrama como un archivo, imprimirlo, normalizar los datos (en caso que usted todavía no lo haya hecho), cambiar el tamaño y la forma de los puntos, y buscar ciertos términos en los genes basándose en la lista de genes de la cual estos genes son derivados.

# Create Table (Crear Tabla)

Esta opción es exclusiva de la Herramienta MAGIC y crea una tabla dinámica. El color predeterminado es una tabla a escala gris, pero usted puede cambiar esto a una escala rojo-verde si usted así lo prefiere. La característica más interesante de esta tabla interactiva es la barra de escala y las tres lengüetas deslizantes. Imagine un set de genes que tiene un gen con una proporción muy grande (por ejemplo +16) y un gen con una proporción muy baja (-16) pero con la mayoría de genes teniendo proporciones entre +3 y -3. Debido a estos dos genes extremos, las diferencias en color en los genes restantes se perderían. Sin embargo, si usted ajusta las lengüetas, usted puede comprimir la escala de color en los extremos y aportar más variación de



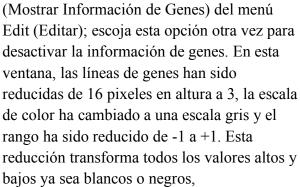
color al rango medio de las proporciones, donde la mayoría de sus genes están localizados. Usted puede utilizar el ratón para arrastrar las lengüetas, o escribir valores numéricos en las casillas correspondientes a cada lengüeta para cambiar los colores. Usted puede escoger ver la información de gen asociada con los genes en el grupo al escoger la opción "Show Gene Info"

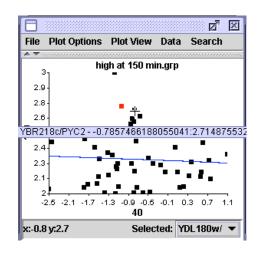


respectivamente, pero permite que los valores intermedios estén en la escala gris.

#### Two Column Plot

Este diagrama le permite seleccionar dos columnas de datos y comparar sus proporciones. Como usted puede ver, algunas comparaciones son más similares que otras. En este diagrama, usted puede seleccionar genes

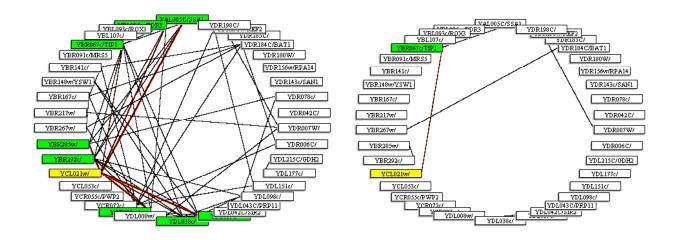




individuales o múltiples (al mantener presionada la tecla Shift mientras usted hace clic). Si usted mueve el ratón sobre un gen, la pantalla le dirá las dos proporciones correspondientes a los dos períodos de tiempo. Usted puede también ver una aproximación en la esquina inferior izquierda.

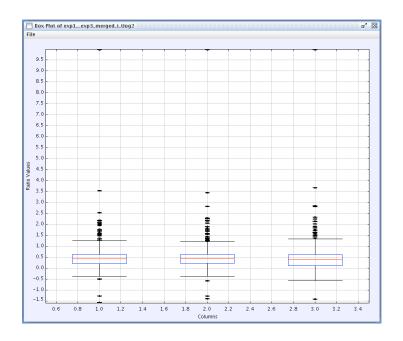
## Circular Display

Otra característica exclusiva de la Herramienta MAGIC es la pantalla circular. Imaginemos que usted ha creado un grupo de genes y que quiere saber la correlación entre coeficientes para estos genes, y en cuáles genes la correlación existe. El ajuste predeterminado es la correlación de coeficientes de 0.8 la cual se indica a su izquierda. Utilizando el menú Display, usted puede cambiar el radio del círculo y el límite para reportar las correlaciones. Cambie el límite a 0.1 (correlación de 0.9) y usted observará menos líneas conectando los genes (en la derecha). En este caso, el mismo gen (amarillo) ha sido seleccionado y los genes que han cumplido con el límite están coloreados en verde con las líneas coloreadas de rojo.



# Box Plot

Usted también puede crear una "box plot" (gráfica de casilleros) estándar la cual muestra el mínimo, cuartil menor, la media, el cuartil mayor y la máxima en formato gráfico. Cuando usted escoja el botón "Box Plot", una "box plot" de todos los genes seleccionados de todas las columna aparecerá, donde cada columna de datos aparecerá como una columna en su "box plot". La caja en sí demuestra los cuartiles mayores y menores y la línea roja la mediana. Las líneas horizontal en la parte posterior e inferior representan el siguiente punto después de 1.5 veces la distancia entre los 25vo y 75vo percentiles de la mediana. Los puntos sobresalientes indican las posiciones de los datos sobresalientes. Un "box plot" le permitirá visualizar experimentos entre columnas. Esto es especialmente útil si usted ha creado réplicas biológicas o más de una chip del mismo experimento.



# Menú "Cluster" (Menú de Agrupación)

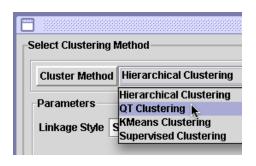
# "Compute..." (Computar) (Control C)

Una vez que usted ha creado un archivo de disimilitud, usted puede agrupar sus datos. Para hacer esto usted debe computar la agrupación utilizando uno de los cuatro métodos. Detalles para estos cuatro métodos se encuentran en la Guía del Instructor.



#### Hierarchical Clustering (Agrupación por Jerarquías)

La agrupación por jerarquías produce una estructura ramificada (un dendograma) al conectar los genes según su similitud de sus datos de expresión. Cuando un gen se une con otro o a un grupo de genes en el árbol, toda la colección de genes es representada como un pseudo-gen individual. La similitud entre un específico gen y el gen (o pseudo-gen) al cual es conectado, es indicada por la longitud horizontal de



las ramas que se unen a él. En cada paso en el algoritmo, los dos genes más similares o pseudogenes son conectados. El proceso continúa hasta que todos los genes se han unido al árbol.

# QT Clustering (Agrupación QT)

La agrupación QT toma cada gen en consideración uno a la vez, construye una agrupación temporaria para cada gen con un valor límite definido por el usuario para encontrar la similitud. Cualquiera que sea el gen que ha conseguido más genes en su grupo es utilizado para crear la agrupación permanente y todos los genes asociados en esta agrupación son eliminados de la lista de genes para la siguiente ronda de creación de agrupaciones permanentes. La agrupación QT repite el proceso de crear agrupaciones temporarias, un gen a la vez, y después forma la segunda agrupación permanente utilizando la agrupación temporaria más grande. Este proceso se repite hasta que todos los genes se encuentren en agrupaciones, o los genes restantes formen agrupaciones más pequeñas que el tamaño definido por el usuario. Estos genes restantes (llamados "singletons") no son presentados en las pantallas de agrupaciones a menos que el usuario haya definido "1" como valor mínimo para una agrupación permanente.

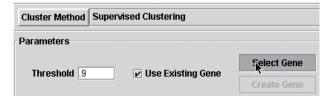
Cuando usted utilice una agrupación QT, usted deberá ajustar el valor límite. El valor predeterminado de 0.9 se traduce en coeficientes de correlación de +0.1 hasta +1.0. Si usted cambia el ajuste del límite a +0.2, usted agrupará genes solamente si sus coeficientes de correlación son de +0.8 a +1.0. El rango de ajuste para el límite va de 0 (correlación de +1.0) hasta 1 (correlación de 0, es decir, no es nada similar) hasta 2 (correlación de -1.0; en dirección opuesta uno del otro). Por esto, al establecer el límite en 2, usted obtendría cada gen en una sola agrupación.

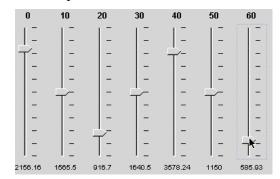
# K-Means Clustering (Agrupación por Método K)

En este método, usted determina *a priori* cuántas agrupaciones usted quiere obtener (donde K es igual al número de agrupaciones) y la Herramienta MAGIC se asegurará que todos los genes se encuentren incluidos en este número de agrupaciones. Éste es el primer paso en los "Self Organized Maps" (Mapas Automáticamente Organizados) pero ambos métodos comienzan con el investigador siendo quien determine cuántas agrupaciones serán generadas.

# Supervised Clustering (Agrupación Supervisada)

Este método ejecuta una agrupación QT pero usted puede definir el límite y escoger un gen alrededor del cual usted quiere construir su agrupación. Esto le permite concentrar su investigación en su gen preferido. En la izquierda, usted puede ver cómo la opción "Use Existing Gene" (Utilizar Gen Existente) está seleccionada. Haga clic en el botón "Select Gene" y después escoja entre los genes en su lista de genes en el archivo de expresión actual.



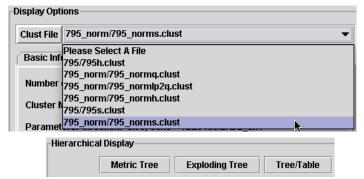


Alternativamente, usted puede deseleccionar la opción "Use Existing Gene" (Utilizar Gen Existente) y después hacer clic en "Create Gene" (Crear Gen). Esto produce una ventana que le permite manipular los deslizadores para crear un perfil de expresión en el cual usted quiera encontrar genes con perfiles similares (basado en el límite que usted escoja). Ésta es una manera fácil de encontrar patrones complejos que le interesen.

# "Display..." (Mostrar...)

Una vez que usted haya creado una agrupación o dos, usted puede visualizarlas en su pantalla. Primero, escoja el archivo de agrupación que usted quiera mirar. Cada tipo de agrupación tiene

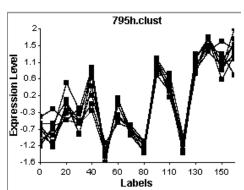
sus diferentes opciones de presentarse.

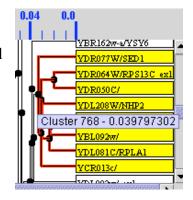


Hierarchical Cluster Display (Mostrado por Agrupaciones de Jerarquía)

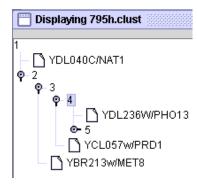
Usted tiene tres opciones para presentar sus agrupaciones, y cada una tiene sus propias opciones. Metric Tree (Árbol Métrico) es exclusivo a la agrupación jerárquica. Este método produce un dendograma con nódulos graficados en límites establecidos. Mientras más pequeño sea el valor del límite, el coeficiente de correlación será más grande. Usted

puede hacer clic en un punto de la ramificación y resaltar todos los genes en esta agrupación como se indica a la derecha. Si su mueve el cursor sobre el punto de ramificación, usted podrá ver el límite exacto el cuál es 1 menos el coeficiente de la correlación (-0.96). Usted puede graficar esta agrupación, y como usted podrá imaginarse, con un coeficiente de correlación tan alto, el diagrama de los puntos normalizados es un grupo muy apretado. Exploding Tree (Árbol de





Expansión) es una forma eficiente de mostrar agrupaciones y gradualmente expandir los contenidos de cada nódulo. En este ejemplo, hay un gen y después todos los otros genes están dentro del segundo nódulo. Mientras usted va haciendo clic en los nódulos, éstos se expanden, y si usted hace clic una segunda vez, se comprimen. Usted puede expandir el nódulo completamente al resaltar el número y haciendo clic en el botón "Explode" (Explotar), o expandir uno a la vez al hacer clic en el nódulo directamente. Usted también puede graficar cualquier agrupación



dentro de un nódulo al hacer clic en el botón "Plot Node As Group" (Graficar Nódulo Como Grupo).

<u>Tree/Table</u> (Árbol/Tabla) es una forma de combinar la tabla de visualización y el dendograma. El dendograma se encuentra al extremo izquierdo y la tabla coloreada (que ocupa la mayoría de la ventana) se indica a la derecha (diagrama que no se indica aquí).

QT Cluster Display (Pantalla de Agrupación QT)

QTClust Display

List Exploding Tree Tree/Table

La Agrupación QT también permite el árbol de expansión y la combinación Tree/Table, pero ha remplazado el Árbol Métrico con la "<u>List</u>" (Lista). La lista le permite ver el número de la raíz del gen para cada agrupación. **Si** usted hace clic en la raíz del gen, entonces todos los genes en esta

agrupación serán identificados. Usted puede graficar esta agrupación como se indica aquí.

Supervised (QT) Cluster Display (Pantalla de Agrupación OT Supervisada)

La agrupación supervisada tiene las mismas opciones de presentar que la agrupación QT. Sin embargo, cuando usted esté escogiendo el tipo de presentación en su pantalla, usted deberá notar que la casilla que indica qué limite ha sido utilizada y cuál gen ha sido utilizado como la raíz. En este caso, ERD2, el exon 1 del receptor KDEL ha

Parameters: threshold=0.95, Gene = YBL040c/ERD2\_ex1

List

sido utilizado como la raíz para esta agrupación con un coeficiente de correlación de 0.95 (gráfica que no se indica).

K-means Cluster Display (Pantalla de Agrupación por Método K)

Las tres opciones de presentación para el método K han sido descritas previamente en esta sección.

\*\*MeansClust Display\*\*

**Exploding Tree** 

#### Crear Dendograma con JTreeView

Cuando las listas de genes se vuelven más largas de 5000 genes, presentar las agrupaciones se vuelve un proceso un tanto lento en la Herramienta MAGIC. Una forma de resolver esta situación es exportar una agrupación computada por la Herramienta MAGIC para que ésta sea vista desde otro programa (software). Nosotros exportamos archivos que pueden ser leídos con el software "open source" Java TreeView. Solamente archivos creados el método de agrupación por jerarquías trabajan por el momento con Java TreeView. Cuando usted haga clic en el botón Export en el diálogo "JTreeView Export Information" (Exportar Información JTreeView), los archivos requeridos para visualizar la agrupación en JTreeView son creados. JTreeView es automáticamente iniciado en su computadora, los archivos son cargados, y un dendograma aparecerá. Usted también puede visualizar los datos en un "karyoscope" el cual puede ayudar a detectar la "aneuploidy"; para hacer esto, reabra el archivo en JTreeView (haga clic en el menú File, Open... (Archivo, Abrir...) y después escoja el archivo que usted viene de exportar y haga clic en Open), después escoja "Karyoscope" del menú "Analysis" (Análisis).

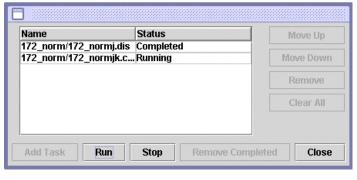
Para más información acerca de Java TreeView, visite el sitio web: <a href="http://jtreeview.sourceforge.net">http://jtreeview.sourceforge.net</a>

#### Menú de Tareas

A medida que sus sets de datos se vuelven más grandes, el tiempo que se requerirá para llevar a cabo todos los cálculos necesarios incrementará rápidamente. Debido a esto, la Herramienta MAGIC le permitirá establecer una lista de tareas que será ejecutada en orden. Usted puede



decirle a la Herramienta MAGIC que usted quiere empezar una serie de pasos y después dejar su ordenador trabajando. La Herramienta MAGIC ejecutará esta secuencia de tareas mientras usted hace otras cosas. Por ejemplo, usted puede establecer una lista de tareas para llevarse a cabo e irse a su casa a descansar. Cuando usted regrese la mañana siguiente,



se dará cuenta la Herramienta MAGIC habrá completado toda la serie de tareas. En este momento, las únicas tareas que pueden ser ejecutadas son los cálculos de disimilitudes y las agrupaciones.

# Task Manager (Gerente de Tareas) (Control Shift M)

La ventana que usted ve en la parte posterior de esta página corresponde al Gerente de Treas. Dicha ventana le permite agregar o eliminar una tarea, cambiar el orden de una tarea, así como ordenar diferentes tareas de mantenimiento.

# Add Task (Agregar Tarea) (Control T)

Esta opción le permite agregar una tarea sin pasar por el Gerente de Tareas.

# Help (Ayuda) (Control H)

Este comando muestra una versión modificada de esta misma Guía del Usuario en la Herramienta MAGIC.

#### **Créditos**

La Herramienta MAGIC versión 1.0 ha sido escrita en JAVA por Adam Abele, Brian Akin, Danielle Choi, Paul Karnik y David Moskowitz. Los contribuidores a las versiones subsecuentes son Mackenzie Cowell, Gavin Taylor, Bill Hatfield, Nicholas Dovidio y Micheal Gordon. Laurie J. Heyer y A. Malcolm Campbell son los consejeros del equipo diseñador de los códigos de este programa. La Herramienta MAGIC ha sido desarrollada en Davidson College y apoyada por la Fundación Nacional de Ciencias (National Science Foundation, NSF), the Duke Endowment, y Davidson College.

Partes del código han sido escritas por Alok Saldanha (JTreeview), The MathWorks y NIST (JAMA matrix library) y Jari Häkkinen y Nicklas Nordborg (BASE). Estas secciones están licenciadas en la GNU Public License Version 2 o en licencias compatibles.

Agradecemos el proyecto Open Source Physics, en particular a Wolfgang Christian y Mario Belloni, por compartir su sabiduría y recursos con nosotros.