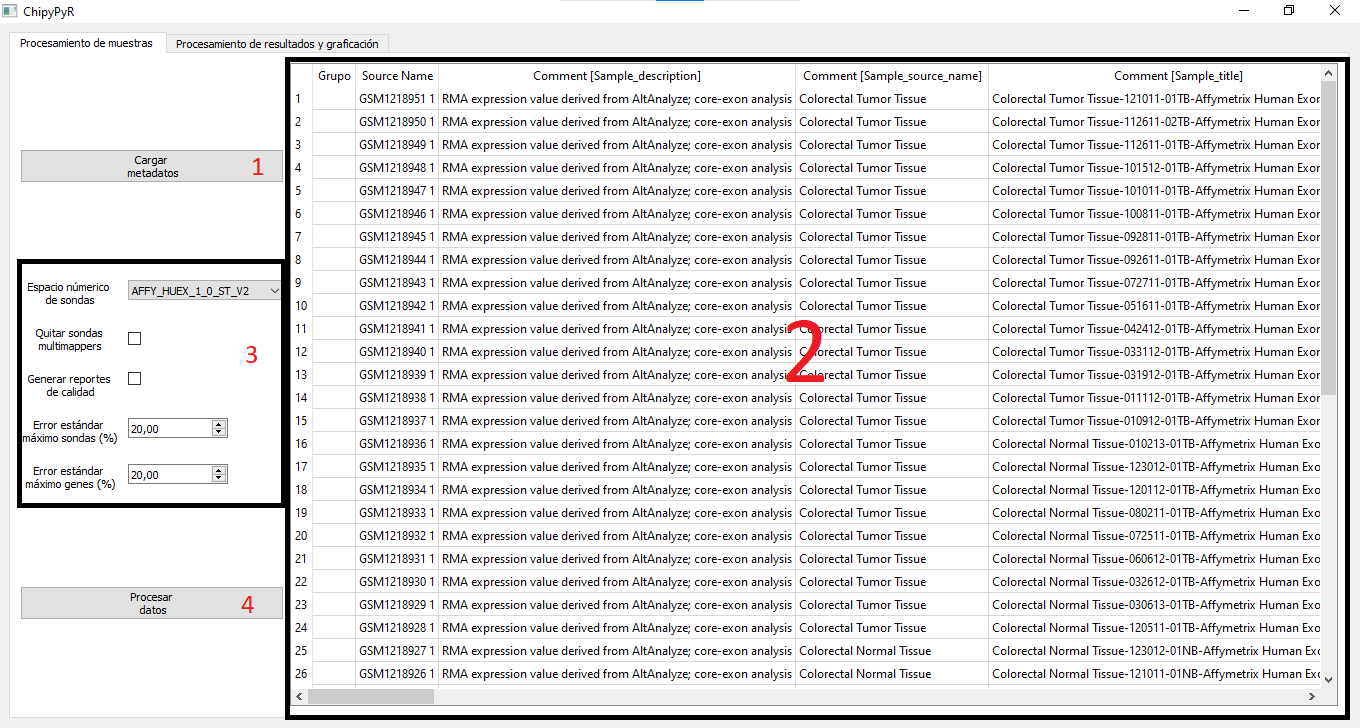
## IMPORTANTE

El único requisito para el software es tener instalado [R](https://cran.r-project.org/bin/windows/base/), el software además supone que se encuentra instalado en la ruta por defecto que es “C:\Program Files\R\R-4.3.1”, para la versión 4.3.1.

## Inicio rápido

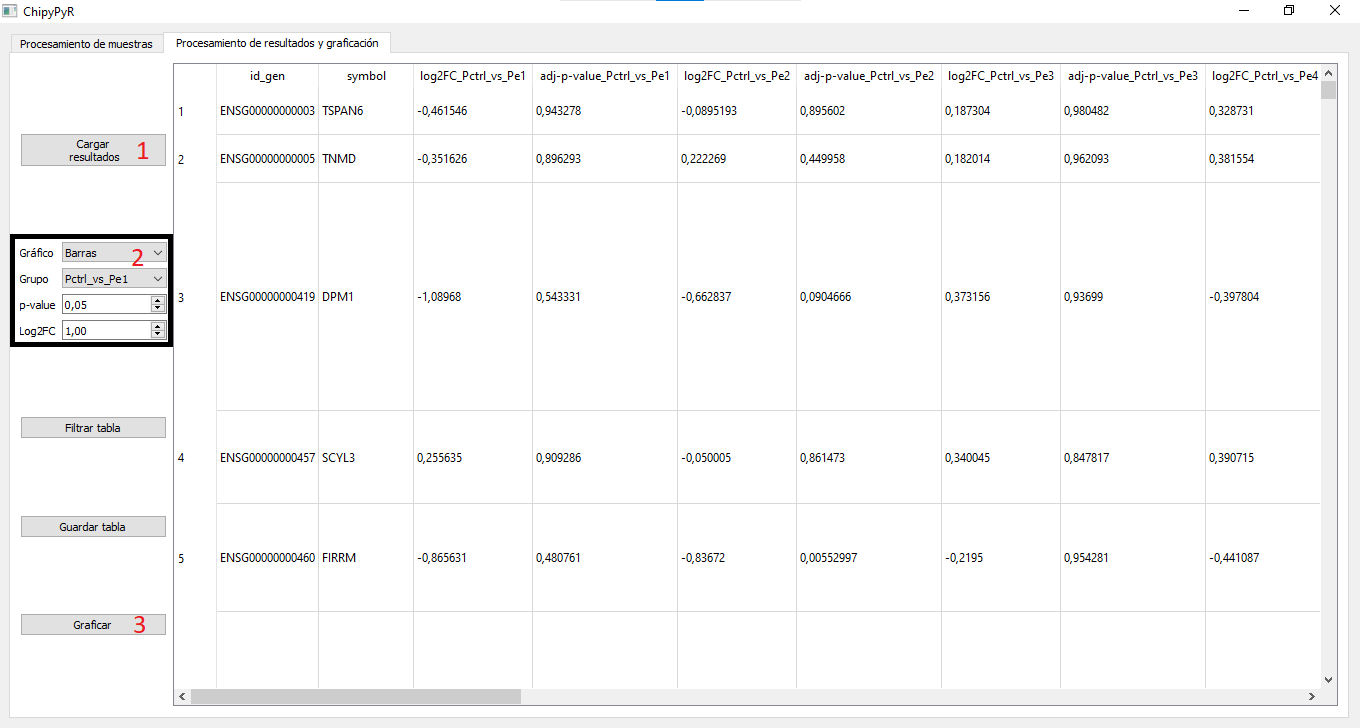
### Procesamiento de muestras



Los pasos a seguir para la generación de resultados son los siguientes:

1. Cargar el archivo de metadatos en formato MAGETAB de las muestras.
2. Una vez cargados los datos aparecerá una columna adicional llamada “grupo” al comienzo, en la cual se debe ingresar un grupo para cada muestra. ***ver controles de tablas***
3. Modificar según sea necesario los parámetros del procesamiento. ***ver procesamiento***
4. Presionar el botón de procesar muestras y elegir la carpeta donde están las muestras, que también será donde se generarán los resultados, y cual de los grupos considerar como grupo de control.

### Procesamiento de resultados y graficación



Los pasos para generar gráficos en la segunda pestaña son los siguientes:

1. Se debe cargar al menos un archivo de resultados generado por la primera parte.
2. Se ajustan los parámetros según sea necesario para el tipo de gráfico que se desea hacer. ***ver proceso de generación de gráficos***
3. Presionar el botón graficar generará una nueva ventana que contendrá el gráfico seleccionado.

\*Si se desea se puede filtrar la tabla en base al p-value y log2FC y luego guardar la tabla actual con esos genes filtrados.

## Formato de entrada y búsqueda de archivos

Todos los archivos manejados por el software deben estar separados por tabulaciones (\t).

El software espera un formato de entrada para las muestras el cual contenga en la primera columna el id de la sonda y en la segunda el valor de intensidad, si contiene más columnas estas se ignoran.

Al indicar el directorio de las muestras, el software las buscará sin tener en cuenta mayúsculas y minúsculas e intenta encontrar algún archivo el cuál contenga en su nombre el valor que se encuentra en la columna “Source name”.

Los archivos de resultados pueden modificarse dentro de ciertas restricciones, para ser aceptados y reconocidos por el software los valores de log2FC y p-value entre los contrastes deben tener dentro del nombre de la cabecera de la columna la secuencia “vs”, y “adj-p-value” o “log2FC”, según corresponda.

## Controles de tablas

Ordenar filas: Al clicar sobre una cabecera de columna los datos se ordenarán según esa columna de manera ascendente o descendente.

Selección de celdas: Se puede seleccionar una celda haciendo clic sobre ella, también se pueden seleccionar múltiples celdas manteniendo presionado control y clicar sobre múltiples celdas, otra opción es seleccionar filas arrastrando desde una hasta otra, un resultado similar puede conseguir clickeando, con la tecla shift apretada, una segunda celda.

Reemplazar múltiples valores: Para simplificar tareas de modificación de celdas se pueden seleccionar múltiples celdas y al presionar la tecla enter se pueden reemplazar los valores de todas esas celdas por uno a elección. Su principal uso está en la asignación de grupos.

Borrar filas: Presionando las teclas suprimir o retroceso se borran todas las filas correspondientes a las celdas seleccionadas.

Mover columnas: Se pueden mover las columnas de las tablas cuyas celdas se hayan seleccionado utilizando las flechas izquierda y derecha o las teclas a y d para mover hacia la izquierda y derecha, respectivamente.

## Procesamiento

El procesamiento aplicado a los datos puede verse de manera detallada en la documentación anexa del software, básicamente consta de los siguientes pasos:

1. Se cargan las muestras según el grupo asignado.
2. Se procesan y filtran las sondas que tengan un error intragrupo mayor al establecido.
3. Se mapean las sondas con sus respectivos genes y se filtran las que no se puedan mapear con ningún gen, y las que mapean a múltiples genes, de estar configurado. En caso de que se hallen sondas numéricas, se considera el espacio numérico que se encuentre seleccionado en los parámetros.
4. Se agrupan los valores de sondas que mapean al mismo gen utilizando la mediana.
5. Se vuelven a filtrar los genes que un mayor error intragrupo al establecido.
6. Se calculan los valores de expresión y p-value para cada gen de los grupos contra el grupo establecido como control.
7. Se enriquecen los resultados con el símbolo, términos GO y vías asociadas al gen.
8. Al finalizar se generará un aviso que también contendrá la cantidad de datos filtrados por etapa.

De estar configurado, después del paso 2 y 6 el software generará una ventana con gráficos que permitan realizar un control de calidad a medida que se va ejecutando el proceso.

## Archivos generados en la carpeta resultados

datos\_limpios.tsv: Archivo que contiene en la primera columna el ID de cada gen a los cuales se encontró alguna sonda mapeada, y en las demás columnas los valores de intensidad calculados para cada muestra, las muestras se encuentran ordenadas en base a los grupos elegidos.

genes\_alto\_error.tsv: Similar al archivo anterior, tiene el ID de cada gen y en cada columna el valor de cada muestra, sin embargo en este caso se genera un archivo diferente con las muestras de cada grupo, además del promedio de expresión del gen y el error estándar.

sondas\_alto\_error.tsv: Idéntico al archivo anterior, sólo que contiene en cambio los id de sonda y es generado al comienzo del procesamiento, en el primer filtrado.

resultado\_final.tsv: Este archivo contiene los resultados del procesamiento que son utilizados en el segundo módulo, contiene ID de los genes, símbolo, datos adicionales de enriquecimiento y los resultados de log2fc y p-value entre cada grupo y el control.

## Proceso de creación de gráficos

Barras: Se quitan las filas de genes donde todos los valores de p-value sean mayores o los valores de log2fc sean menores en valor absoluto a los parámetros establecidos, luego se separan los datos según la cantidad en distintas pestañas y se muestran los valores de expresión para cada grupo en distintos colores, se utiliza como etiqueta el símbolo del gen, de estar disponible, de lo contrario se utiliza el ID.

Expression Atlas: Se muestra un selector de opción para elegir a la página de qué gen de los que se encuentran en el archivo de resultados se quiere acceder en Expression Atlas.

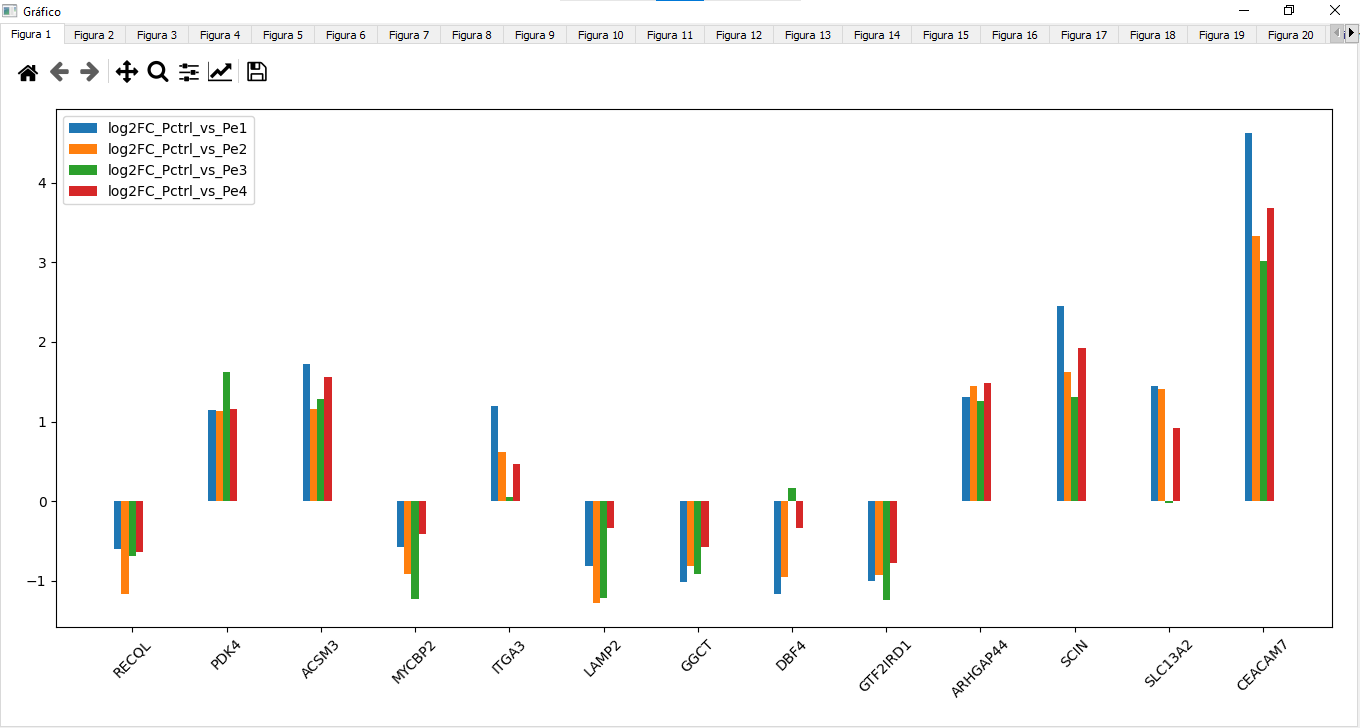
Heatmap: El procesamiento es idéntico al de barras.

Reactome: Se toman los datos del grupo seleccionado en los parámetros y se filtran todos los genes cuyos p-values sean mayores al establecido o si el valor de log2fc es menor en valor absoluto, los genes que pasen el filtro son utilizados para una consulta con Reactome y se muestra un gráfico Fireworks.

Venn: Se filtran en cada grupo los genes cuyos p-values sean mayores al establecido o si el valor de log2fc es menor en valor absoluto, luego se contabiliza cuantos se comparten en cada intersección de los grupos y se los muestra en un diagrama de Venn.

Volcán: Se toman los valores de p-value y log2fc para cada gen del grupo seleccionado y se colorean los puntos en el gráfico según los valores de corte establecidos en los parámetros antes de graficar.

## Ventana de gráficos



Las ventanas de gráficos contienen pestañas en la parte superior que permite cambiar entre las distintas figuras generadas si los datos son demasiados o si se generaron múltiples gráficos, también cada figura cuenta con un panel de opciones que permiten modificar la visualización moviendo la vista, haciendo zoom, modificando márgenes y también permite guardar la imagen. Cabe recalcar que el orden de los genes en gráficos como el de barras, mostrado en el ejemplo, depende del orden de la tabla de resultados.

## Errores conocidos

### El software se cierra durante el cálculo de expresión sin ningún error

Si el usuario no tiene permisos de escritura en la carpeta donde se encuentra R, más específicamente, en la carpeta “library”, el programa puede cerrarse inesperadamente mientras se está realizando el cálculo de expresión, ya que el software no puede instalar y actualizar los paquetes de R necesarios.

