

# Introducción al análisis de datos de microarrays

## ***Ejercicio (1): Detección de genes diferencialmente expresados***

### **Objetivos**

- Este ejercicio pretende realizar una introducción al análisis de microarrays mediante un estudio sencillo de expresión diferencial.
- El programa utilizado será **BRB-Array-Tools** aunque puede adaptarse fácilmente para su uso con otras herramientas.

### **Objetivos del estudio**

- Este estudio pretende hallar genes diferencialmente expresados antes y después de la quimioterapia
- Para ello analizaremos datos obtenidos de biopsias de pacientes de cáncer de pecho, **antes** y **después** de un tratamiento de 16 semanas con *doxorubicina*
- Se basa en el trabajo de Korn et. al. *Identifying Pre-Post Chemotherapy Differences in gene Expression in Breast Tumors*.

### **Diseño el estudio**

- Se utilizaron muestras de RNA de 20 pacientes de cáncer de pulmón, que se hibridaron en arrays de cDNA.
- Se utilizó un diseño de referencia comparando cada muestra con una mezcla de RNA de 11 líneas celulares. (*¿Que otro diseño habría sido posible o razonable?* ).
- Los datos están apareados, aunque en arrays distintos: Hay dos muestra por paciente una antes y otra después de la quimioterapia.

### **Flujo de trabajo del estudio**

Como en todo estudio de microarrays procederemos en una serie de pasos ordenados que de forma simplificada reproducen el procedimiento real

1. Cargar los datos.
2. Filtrar los puntos defectuosos y ajustar las intensidades
3. Normalización y verificación.
4. Realización de pruebas estadísticas.
5. Interpretación de los resultados

#### **1. Entrada/Carga de los datos**

- Arranca el programa y carga los datos de la fuente que se indique (PEROU.XLS o bien PEROU.TXT)
- Observa los datos, las anotaciones y los descriptores de los grupos.

## 2. Filtrado de puntos defectuosos

- Vamos a filtrar los datos de forma que los puntos cuya intensidad no alcance un valor de 100 serán excluidos del análisis
- También eliminaremos las intensidades marcadas con un “flag” distinto a cero.

*No te preocupes si esto parece arbitrario, porque lo es. Se trata de reglas empíricas obtenidas de la experiencia de análisis anteriores para excluir valores que contribuyen, más que nada, con ruido.*

## 3. Normalización

- Una rápida inspección de los datos permite decidir si hace falta normalizarlos.
- En primer lugar centraremos los datos substrayendo la mediana del array a todos sus valores.
- A continuación realizaremos una transformación no lineal (loess)
- Un MA plot posterior a la normalización no servirá para comprobar que tal ha funcionado el proceso

## 4. Comparación de grupos

- La detección de genes diferencialmente expresados puede hacerse:
  - de forma visual (*tan solo con fines exploratorios!!!*)
  - mediante los tests estadísticos adecuados
- Un diagrama de dispersión puede dar una idea de que genes se encuentran diferencialmente expresados
- Tras la exploración inicial la mejor aproximación es un test estadístico
  - Permite estimar la significación de los resultados
  - Permite controlar el problema de la multiplicidad de tests

## 5. Interpretación de los resultados

- Observa la lista de genes seleccionados. Puedes relacionar alguno de ellos con el cáncer?
- Una ayuda para la interpretación de estos resultados puede obtenerse a través de la Gene Ontology
  - Busca que clases de la GO se encuentran enriquecidas entre la lista de genes seleccionados.
  - Encuentras algunas que puedas relacionar con la enfermedad?