

## Tarea 5 – Coexpresión y medidas de centralidad

### Objetivos

1. Entender el concepto de redes de coexpresión
2. Calcular medidas de centralidad en grafos
3. Visualizar grafos con Cytoscape
4. Practicar programación
5. Utilizar WGCNA para producir redes de coexpresión y clustering
6. Utilizar Cytoscape para calcular medidas de centralidad
7. Inferir hipótesis biológicas a partir de análisis de redes y enriquecimiento funcional

**Datos a trabajar:** Se incluye con esta tarea un conjunto de datos de microarreglos correspondiente a un ensayo de expresión de genes en hígado de ratón femenino. Los datos provienen de un experimento realizado en hígados de ratones que pretendía hallar la relación entre las tendencias de expresión de varios genes y marcadores génicos con el peso de los individuos y otros indicadores fenotípicos relacionados. Los datos de expresión y de fenotipos contienen muestras para cada uno de los ratones incluidos en el experimento, cuyo ID individual comienza con el prefijo F2.

### Pasos a seguir

#### **Parte I**

1. Descargar Rstudio y R de ser necesario. Cargar el script de WGCNA y correr las líneas para instalar los paquetes necesarios al principio del mismo. Descargar los datasets y mirar qué clase de datos incluye cada uno. Esto es importante para entender los cambios que se deben hacer en el script.
2. Cambiar los parámetros indicados como PUNTOS en el script, de acuerdo a los resultados que vaya arrojando el análisis. Deben ir corriendo el código con cuidado de leer los comentarios para entender lo que se está haciendo y donde se deben cambiar parámetros. Varios de estos dependen de analizar las gráficas que se vayan produciendo. (Para correr los bloques de código en Rstudio se debe seleccionar las líneas que se quieren ejecutar y orpimir shift + Click izquierdo).
3. Cargar en Cytoscape el archivo de ejes que se genera en la última parte del script. Utilizar la herramienta analyze para obtener las medidas de centralidad. Adjuntar las gráficas generadas y analizar qué pueden indicar estas medidas con respecto a la red del módulo escogido y qué relevancia a nivel de correlación por función puede tener con los resultados obtenidos del enriquecimiento (Para esto, se ha generado un archivo previamente). Generar una hipótesis biológica que relacione todos estos hallazgos.

Puede también investigar la función específica de algunos de los genes que se agruparon en el cluster escogido y relacionarlo.

## **Parte II**

1. Implementar una función que calcule la matriz de coexpresión dado un archivo con información de niveles de expresión, como el de los datos de ejemplo. Ignorar las columnas que no contienen información de expresión. Una vez cargados los datos del archivo, se recomienda remover los genes que no reporten niveles de expresión en ninguno de los ensayos.

2. Implementar una función que dada una matriz de coexpresión y un valor mínimo de correlación, retorne un grafo no dirigido que represente una red de coexpresión. En este grafo, dos proteínas deben estar conectadas si su coeficiente de coexpresión es mayor o igual que el valor mínimo dado por parámetro. El grafo se puede implementar como matriz de adyacencias o como lista de adyacencias.

3. Desarrollar funciones que reciban la red PPI y generen la siguiente información:

- Grado de cada vértice
- Distribución de grados
- Densidad total
- Coeficiente de agrupamiento de cada vértice
- Espectro de agrupamiento

4. Generar un archivo de texto con la información de la red para poderla visualizar en cytoscape. El archivo debe tener al menos dos columnas:

- Id gen 1
- Id gen 2

Opcionalmente pueden incluir una columna con el coeficiente de correlación entre cada par de genes.

Una vez sea generado este archivo, cargarlo en Cytoscape para poderlo visualizar.

**Bono. Simulación de redes.** Hacer una función que simule una red aleatoria con los mismos vértices de la red de coexpresión y con una cantidad de ejes aleatorios igual a la cantidad de ejes de la red. Calcular sus medidas de centralidad y explicar las semejanzas y diferencias con las medidas de la red de coexpresión.